Goodness-of-Fit Test for Monotone Functions

CÉCILE DUROT
Département de Mathématiques, Université Paris-Sud

LAURENCE REBOUL
Département de Mathématiques, Université de Poitiers

ABSTRACT. In this article, we develop a test for the null hypothesis that a real-valued function belongs to a given parametric set against the non-parametric alternative that it is monotone, say decreasing. The method is described in a general model that covers the monotone density model, the monotone regression and the right-censoring model with monotone hazard rate. The criterion for testing is an $L_p$-distance between a Grenander-type non-parametric estimator and a parametric estimator computed under the null hypothesis. A normalized version of this distance is shown to have an asymptotic normal distribution under the null, whence a test can be developed. Moreover, a bootstrap procedure is shown to be consistent to calibrate the test.

Key words: bootstrap, composite hypothesis, least concave majorant, monotone density, monotone hazard rate, monotone regression, non-parametric alternatives

1. Introduction

Monotonicity is a shape restriction that emerges naturally in many application areas. In economics for instance, empirical evidence often suggests a monotonic relationship between a covariate and a response variable, as it is the case of the presumed decreasing (respectively increasing) relationship between price and demand (respectively supply). In reliability or biological studies, the behaviour of a system or human lifetime on a specific time period can often be appropriately described by a distribution with a monotone hazard rate. In such situations, one may hope to obtain more accurate modelling by using statistical methods specifically designed for monotone functions.

Consider the problem of estimating a function $\lambda$ under the constraint that it is monotone, say decreasing (the increasing case can be treated likewise). Adopting a non-parametric approach, one can build in various models a non-parametric estimator under the monotonicity constraint. Alternatively, one can postulate a parametric model that respects the monotonicity constraint, and thus consider a parametric estimator. In that case, the presumed model should be checked by an appropriate goodness-of-fit test. It is the aim of this article to build such a test: we aim to test that $\lambda$ belongs to a given parametric family that respects the monotonicity constraint, against the alternative that $\lambda$ is decreasing.

A pioneer paper about non-parametric inference under monotonicity constraint is that of Grenander (1956), where it is shown that the non-parametric maximum likelihood estimator of a non-increasing density, based on an i.i.d. sample, is the slope of the least concave majorant of the empirical distribution function. Then, Brunk (1970) proves that the non-parametric least-squares estimator of a non-increasing regression mean is a Grenander-type estimator, in the sense that it is the slope of the least concave majorant of an estimator of the primitive of the function to estimate. Huang & Wellner (1995) study a Grenander-type estimator for a monotone hazard rate in a right-censoring model. Reboul (2005) and Durot (2007) study such estimators in a general model that covers the cases of density, regression
and right-censoring models: Reboul (2005) provides a non-asymptotic control for the $L_1$-risk and proves that the estimator is spatially adaptive, whereas Durot (2007) gives the limit distribution of the $L_p$-error as the number of observations tends to infinity.

On the problem of testing a parametric null hypothesis against a non-parametric alternative, many papers are inspired by the smooth test of Neyman (1937); see among others Eubank & Hart (1992), Ledwina (1994), Fan & Huang (2001) Fan (1996), Aerts et al. (1999), Gray & Pierce (1985) and Peña (1998). There, the alternative is modelled as the product between the null parametric function and a series expansion along an orthonormal family of functions. The number of selected functions in the expansion turns out to be a smoothing parameter, and testing the null hypothesis reduces to test the nullity of the coefficients in the expansion. This can be done thanks to likelihood ratio or score test statistics with a data-driven calibration of the smoothing parameter; see Ledwina (1994) or Fan (1996). Another popular approach for testing a parametric model versus a non-parametric alternative is to reject the null hypothesis if a non-parametric estimator is too far from a parametric estimator computed under the null hypothesis; see among others Hardle & Mammen (1993), Alcalá et al. (1999), Stute & González-Manteiga (1996), Hart (1997), Liero et al. (1998), Liero & Läuter (2006). In the aforementioned papers, the alternative is not subjected to a monotonicity constraint. If the function under interest is known to be monotone, one can even so use these tests and omit the monotonicity assumption, but it seems more relevant to use statistical methods specifically designed for monotone functions. In this spirit, Durot & Tocquet (2001) carried out a goodness-of-fit test for a simple null hypothesis in a regression model with monotone regression mean: they reject the null hypothesis if the $L_1$-distance between the null hypothesis and the Brunk estimator is too large. Ducharme & Fontez (2004) adapt the smooth test of Neyman to a regression setting where the regression mean is assumed positive and increasing.

In this article, we propose a goodness-of-fit test that generalizes the method of Durot & Tocquet (2001) in two ways. First, we allow composite null hypotheses; second, we address the testing problem in a fairly general setting that covers the density, regression and right-censoring models. The criterion for testing is an $L_p$-distance between a Grenander-type estimator and a parametric estimator under the null hypothesis. A normalized version of this distance is shown to have an asymptotically normal distribution under the null, whence a goodness-of-fit test can be developed. To overcome the difficulty of estimating the parameters on which the asymptotic normal distribution depends, we establish the consistency of a bootstrap procedure to calibrate the test.

The rest of the article is organized as follows. In section 2, we present the test statistic and study its asymptotic distribution under the null hypothesis. Calibrations are described in section 3. In section 4, we detail our method in the aforementioned models. A simulation study is reported in section 5. The proofs are deferred to the Appendix.

2. The test statistic

Assume one observes $X^{(n)}$ the distribution of which depends on a monotone function $\lambda : [0, 1] \to \mathbb{R}$ and possibly on a nuisance parameter $\eta \in H$, where $H$ is a set that could be of infinite dimension. Typically, $X^{(n)}$ consists of $n$ independent but not necessarily identically distributed random variables. The nuisance parameter could be, for instance, the distribution of a censoring variable. Assume that $\lambda$ is defined on a compact interval and without loss of generality, assume (possibly changing scale and origin) that this interval is $[0, 1]$ and (possibly changing $\lambda$ into $-\lambda$) that $\lambda$ is decreasing. We wish to test the parametric null hypothesis.
Goodness-of-fit test for monotone functions

\[ H_0: \lambda \in \{ \lambda_\theta, \theta \in \Theta \} \]  

against the non-parametric alternative that \( \lambda \) is decreasing, where \( \Theta \subset \mathbb{R}' \) is a given set and for every \( \theta, \lambda_\theta \) is a given decreasing function on \([0, 1]\).

The criterion we consider for testing relies on the Grenander-type estimator introduced in Durot (2007). Precisely, we assume in the sequel that we have at hand a cadlag step estimator \( \Lambda_n : [0, 1] \to \mathbb{R} \) of the cumulative function

\[ \Lambda(t) = \int_0^t \lambda(u) \, du, \quad t \in [0, 1] \]

and we define \( \hat{\lambda}_n \) as the left-hand slope of the least concave majorant of \( \Lambda_n \), with \( \hat{\lambda}_n(0) = \lim_{t \to 0} \hat{\lambda}_n(t) \). The Grenander-type estimator \( \hat{\lambda}_n \) can easily be computed thanks to simple algorithms such as the Pool-Adjacent-Violators-Algorithm (PAVA); see Barlow et al. (1972).

It is entirely data-driven and from Reboul (2005), it is known to be spatially adaptive in various models. Our test statistic is

\[ S_{pm} = \int_0^1 |\hat{\lambda}_n(t) - \hat{\lambda}_{\theta_0}(t)|^p \, dt, \quad (2) \]

where \( \hat{\theta}_n \) is a suitable estimator of \( \theta \) under \( H_0 \) and \( p \) is a fixed real. It should be mentioned that in the particular case of a simple null hypothesis (of the form \( H_0: \lambda = \lambda_0 \)), under \( H_0 \), \( S_{pm} \) is the \( L_p \)-error of \( \hat{\lambda}_n \) so its asymptotic distribution is given by theorem 2 of Durot (2007); our main task here is to generalize the method of Durot (2007) to the case of a possibly composite null hypothesis.

We need some notation to describe the asymptotic distribution of \( S_{pm} \) under \( H_0 \). Let \( W \) be a standard two-sided Brownian motion on \( \mathbb{R} \),

\[ X(a) = \arg\max_{a \in \mathbb{R}} \{ -(a - a)^2 + W(a) \}, \quad a \in \mathbb{R}, \]

\[ \mu_p = \mathbb{E}[X(0)^p] \quad \text{and} \quad k_p = \int_0^\infty \text{cov}(|X(0)|^p, |X(a) - a|^p) \, da. \]

Note that \( \mu_p \) and \( k_p \) are both well defined and finite. Let \( \| \cdot \| \) denote the Euclidean norm on \( \mathbb{R}^r \). For every \( \theta \in \Theta \), let \( \mathbb{P}_{\theta_0} \) and \( \mathbb{E}_{\theta_0} \) denote the underlying probability and expectation when \( \lambda = \lambda_0 \), where we recall that \( \eta \) is a nuisance parameter. Note that for typographical convenience, we omit the dependence with respect to \( n \) in the notation. Let \( M_{\eta_0} = \Lambda_n - \Lambda_{\theta_0} \), where \( \Lambda_0(t) = \int_0^t \lambda_0(u) \, du, \ t \in [0, 1] \). Finally, for every \( f_\theta : [0, 1] \to \mathbb{R} \), we set (when it exists)

\[ \hat{f}_\theta(t) = \left( \frac{\partial}{\partial \theta_1} f_\theta(t), \ldots, \frac{\partial}{\partial \theta_r} f_\theta(t) \right), \]

where \( \theta_i \) is the \( i \)-th component of \( \theta \), \( i = 1, \ldots, r \).

We then make the following assumptions.

(A) For every \( \theta \in \Theta \), \( \lambda_\theta \) is decreasing and differentiable on \([0, 1]\) with

\[ \inf_t |\lambda_\theta'(t)| > 0, \]

and there are \( C > 0 \) and \( s \in (3/4, 1] \) such that for all \( t, u \),

\[ |\lambda_\theta'(t) - \lambda_\theta'(u)| \leq C |t - u|^s. \]
(B) Let $B_n$ be either a Brownian bridge or a Brownian motion. For all $\theta$ and $\eta$, there exist $q > 12$, $C > 0$, $L_{0}\eta: [0, 1] \to \mathbb{R}$ and versions of $M_{\eta\theta}$ and $B_n$ such that

$$
P_{0}\eta \left[ n^{-1/4} \sup_{t \in [0,1]} \left| M_{\eta\theta}(t) - n^{-1/2} B_n \circ L_{0}\eta(t) \right| > x \right] \leq C x^{-q}$$

for all $x \in (0, n]$, and

$$
\mathbb{E}_{0}\eta \left[ \sup_{u \in [0,1], x \geq 2 \sup_{t \in [0,1]} \left| t - u \right| \leq x} \left( M_{\eta\theta}(u) - M_{\eta\theta}(t) \right)^2 \right] \leq \frac{C}{n}
$$

for all $x > 0$ and $t \in \{0, 1\}$. Moreover, $L_{0}\eta$ is increasing and twice differentiable on $[0, 1]$ with $L_{0}\eta(0) = 0$, $\sup_{t} |L_{0}\eta(t)| < \infty$ and $\inf_{t} L_{0}\eta(t) > 0$.

(C) $\hat{\lambda}_{n}(0)$ and $\hat{\lambda}_{n}(1)$ are stochastically bounded under $H_0$.

(D) $\hat{\theta}_{n} = \theta + \mathcal{O}(n^{-1/2})$ for every $\theta$ and $\eta$.

(E) For every $t \in [0, 1]$, the functions $\theta \mapsto \hat{\lambda}_{\theta}(t)$ and $\theta \mapsto \hat{\lambda}'_{\theta}(t)$ are differentiable on the convex hull of $\Theta$ and for every $\theta$, the function $t \mapsto \|\hat{\lambda}_{\theta}(t)\|$ is bounded on $[0, 1]$. Moreover, for every $\theta$ and $\varepsilon > 0$ there exists $\delta > 0$ such that

$$
\sup_{\|\theta - \theta'\| \leq \delta} \sup_{t \in [0,1]} \|\lambda_{\theta}(t) - \lambda_{\theta'}(t)\| \leq \varepsilon,
$$

and for every $\theta$, there exist $\delta > 0$ and $\varepsilon > 0$ such that

$$
\sup_{\|\theta - \theta'\| \leq \delta} \sup_{t \in [0,1]} \|\lambda'_{\theta}(t)\| \leq \varepsilon.
$$

Assumptions (A), (B) and (C) are an adaptation to our testing problem of the assumptions in Durot's theorem 2: (A) is a smoothness assumption on $\lambda$ under $H_0$ whereas (B) ensures that $M_{\eta\theta}$ can be approximated in distribution by either a Brownian motion or a Brownian bridge with possibly non-standard variance function. Note that the constants involved in these assumptions may depend on $\theta$ and $\eta$ but for typographical convenience, this does not appear in the notation. Similar to Durot's theorem 2, the assumptions that $s > 3/4$ and $q > 12$ could probably be weakened at the price of more technicalities. A sufficient condition for assumption (C) to hold is given by lemma 1 in Durot (2007). Assumptions (A), (B) and (C) allow computing the asymptotic distribution of the $L_{p}$-error of $\hat{\lambda}_{n}$ under $H_0$, whereas assumptions (D) and (E) are designed to ensure that $\hat{\lambda}_{n}$ properly estimates $\lambda_{\theta}$ under $H_0$.

Under these assumptions, we obtain the asymptotic distribution of the test statistic under $H_0$. Theorem 1

Assume (A)–(E) and let $l_{0}\eta = L_{0}\eta'$. Then for every $p \in [1, 5/2)$, $\theta$ and $\eta$,

$$
n^{1/6} \left( n^{\sigma_3} S_{\eta p} - \mu_p \int_0^1 |4\lambda'_{\theta}(t) l_{0}\eta(t)|^{\rho_3} \, dt \right)
$$

converges in distribution under $P_{0}\eta$ as $n \to \infty$ to the Gaussian law with mean zero and variance

$$
\sigma_p^2 = 8k_p \int_0^1 \left| 4\lambda'_{\theta}(t) l_{0}\eta(t) \right|^{2(p-1/3)} \, l_{0}\eta(t) \, dt.
$$

(3)

Let us comment on the choice of $p$. The constraint $p \in [1, 5/2)$ comes from the fact that for a large $p$, the contribution of the boundaries dominates, which has the effect of changing the asymptotic behaviour of $S_{\eta p}$. The choice $p = 1$ is the simplest one because $\mu_1$ and $8k_1$ are known to approximately equal 0.41 and 0.17, respectively (see Groeneboom, 1985), and the asymptotic variance reduces to $\sigma_1^2 = 8k_1 L_{0}\eta(1)$ in this case. We have observed on a simulation study not reported here that the performances of the test defined with $p = 2$ are quite similar.
to the ones of the test defined with $p = 1$. Thus, hereafter we are mainly concerned with the choice $p = 1$.

To conclude this section let us briefly discuss a possible version of the method in semi-parametric models with covariates. Assume we observe $X^{(n)}$ the distribution of which depends on a decreasing function $\lambda$ and on a nuisance parameter $\eta = (\beta, \gamma)$, where $\beta \in \mathbb{R}^p$ describes the effect of a covariate and $\gamma$ belongs to a possibly infinite dimensional space. For instance, in the Cox model, $X^{(n)}$ is an $n$-sample of the pair $(T, Z)$ where $T$ is a lifetime, $Z \in \mathbb{R}^p$ is a covariate with unknown distribution $\gamma$ and the conditional hazard rate of $T$ given $Z$ is

$$t \mapsto \exp(\beta^T Z) \lambda(t).$$

In typical situations, we have at hand a cadlag step estimator $\hat{\lambda}_n$ of $\lambda$ in the sub-model where $\beta$ is known, an estimator $\hat{\beta}_n$ of $\beta$ which converges at the $\sqrt{n}$-rate to a centred Gaussian distribution and an estimator $\hat{\theta}_n$ which converges to $\theta$ at the $\sqrt{n}$-rate under the null hypothesis (1). In this setting, define $\hat{\Lambda}_n$ to be $\hat{\Lambda}_{n, \hat{\beta}_n}$ and $\hat{\lambda}_n$ to be the Grenander-type estimator based on $\hat{\Lambda}_n$, and define $S_n$ by (2). It is expected that under suitable smoothness assumptions, an analogous assumption to (B) holds so that under $H_0$, $S_n$ is asymptotically Gaussian with a similar asymptotic behaviour as in theorem 1.

3. Calibration

3.1. Plug-in calibration

For the sake of simplicity, we focus here on the case $p = 1$. One obtains a test with prescribed asymptotic level as an immediate consequence of theorem 1, by simply plugging-in suitable estimators of the unknown quantities in the limit distribution.

**Corollary 1**

Let $\alpha \in (0, 1)$ and $q_\alpha$ be the $\alpha$-upper percentile point of the standard Gaussian law. Assume (A)-(E) and let $l_{0\gamma} = L_{0\gamma}$. Let $\hat{l}_n$ and $\hat{L}_n$ be estimators such that for every $0$ and $\eta$,

$$n^{1/6} \sup_{t \in (0,1)} |\hat{l}_n(t) - l_{0\gamma}(t)| \quad \text{and} \quad |\hat{L}_n(1) - L_{0\gamma}(1)|$$

stochastically converge to zero under $P_{0\gamma}$. Then the test with critical region

$$\left\{ n^{1/3} S_{1n} > \mu_1 \int_0^1 |4\lambda'_{0\eta}(t) \hat{\lambda}_n(t)|^{1/3} \, dt + n^{-1/6} q_\alpha \sqrt{8k_1 \hat{L}_n(1)} \right\}$$

has asymptotic level $\alpha$.

In principle, $\hat{l}_n$ and $\hat{L}_n$ could be any estimator satisfying (4). One can consider for instance estimators of the form $\hat{L}_n = \hat{L}_{\hat{\theta}_n, \hat{\eta}_n}$ and $\hat{l}_n = \hat{l}_{\hat{\theta}_n, \hat{\eta}_n}$ where $\hat{\eta}_n$ estimates $\eta$.

3.2. Bootstrap calibration

From the point of view of implementing the test with the aforegiven calibration, one needs to estimate several parameters on which the asymptotic distribution of the test statistic under $H_0$ depends. Besides this difficulty, the Gaussian approximation may be misleading for moderate sample sizes as already mentioned by Durot & Tocquet (2001) in the particular case they consider. To overcome these difficulties, we establish in this section the validity of a bootstrap procedure to calibrate the test.
Hereafter, $P_{\theta \eta}$ denotes the distribution of $X^{(n)}$ under $P_{\theta \eta}$, $P$ denotes the true distribution of $X^{(n)}$, $(X, A)$ denotes the measurable space on which $X^{(n)}$ takes values and $\theta_n$ denotes an estimator of $\theta$. We write $(\hat{\theta}_n, \hat{\eta}_n) = \Phi_n(X^{(n)})$ for a given measurable function $\Phi_n$. Assume that for every $A \in A, x \mapsto P_{\theta \eta}(A)$ is $A$-measurable. Then, we can consider a measurable space $\Omega^\ast$ equipped with a probability measure $Q$, and a measurable pair $(X^{(n)}, X^{(n)^{\ast}}): \Omega^\ast \rightarrow X \times X$, such that for every $B \in A \otimes A$,

$$Q((X^{(n)}, X^{(n)^{\ast}}) \in B) = \int \int 0_B(x, y) \, dP_{\theta \eta}(y) \, dP(x).$$

We call $X^{(n)^{\ast}}$ the bootstrap observation; its conditional distribution given $X^{(n)}$ is $P_{\hat{\theta}_n \hat{\eta}_n}$. As is customary, we denote by $P^\ast$ the probability measure defined on $\Omega^\ast$ by

$$P^\ast((X^{(n)}, X^{(n)^{\ast}}) \in B) = Q((X^{(n)}, X^{(n)^{\ast}}) \in B | X^{(n)})$$

for every $B \in A \otimes A$ and we denote by $E^\ast$ the corresponding expectation.

Let $\Lambda^\ast_n, \hat{\theta}^\ast_n$ and $\hat{\lambda}^\ast_n$ be computed in the same manner as $\Lambda_n, \hat{\theta}_n$ and $\hat{\lambda}_n$ but with $X^{(n)}$ replaced by $X^{(n)^{\ast}}$, and define the bootstrap version of $S_{pn}$ by

$$S^\ast_{pn} = \int_{0}^{1} \left| \hat{\lambda}^\ast_n(t) - \hat{\lambda}^\ast_n(t) \right|^p \, dt.$$  \hspace{1cm} (6)

We propose to reject $H_0$ if $S^\ast_{pn}$ exceeds the $z$-upper percentile point $q^\ast_{pnx}$ of the conditional distribution of $S^\ast_{pn}$. It should be noticed that for practical purposes, $q^\ast_{pnx}$ can be suitably estimated by Monte Carlo simulations. To show that this bootstrap calibration is consistent, we define $M^\ast_n = \Lambda^\ast_n - \Lambda^\ast_n$ and we make the following assumptions, which are bootstrap versions of assumptions (B), (C) and (D):

(B*) Let $B^\ast_n$ be either a Brownian bridge or a Brownian motion under $P^\ast$. For all $\theta$ and $\eta$, with $P_{\theta \eta}$-probability tending to one, there exist a constant $C > 0$, an $X^{(n)}$-measurable $L^\ast_n: [0, 1] \rightarrow \mathbb{R}$ and versions of $M^\ast_n$ and $B^\ast_n$ such that

$$P^\ast \left[ n^{1 - 1/\eta} \sup_{t \in [0, 1]} \left| M^\ast_n(t) - n^{-1/2} B^\ast_n \circ L^\ast_n(t) \right| > x \right] \leq \frac{C x^{-\eta}}{n}$$

for all $x \in (0, n]$, and

$$E^\ast \left[ \sup_{u \in [0, 1], u/2 \leq |t - u| \leq x} \left( M^\ast_n(u) - M^\ast_n(t) \right)^2 \right] \leq \frac{C x}{n}$$

for all $x > 0$ and $t \in \{0, 1\}$. Moreover, $L^\ast_n$ is continuously differentiable with first derivative $L^\prime_n$ and, setting $l^\prime_{\theta \eta} = L^\prime_{\theta \eta}$, there exists $s_0 > 3/4$ such that

$$\sup_{t \in (0, 1)} |l^\prime_{\theta \eta}(t) - L^\prime_n(t)| = O_{P_{\theta \eta}}(n^{-s_0/3}).$$ \hspace{1cm} (7)

(C*) For all $\theta$ and $\eta$, there exists an $X^{(n)}$-measurable set on which $\hat{\lambda}^\ast_n(0)$ and $\hat{\lambda}^\ast_n(1)$ are of the order $O_{P^\ast}(1)$, and which $P_{\theta \eta}$-probability tends to one as $n \rightarrow \infty$.

(D*) For all $\theta$, $\hat{\theta}^\ast_n = \hat{\theta}_n + O_{P^\ast}(n^{-1/2})$ on the set $\{ \| \hat{\theta}_n - \theta \| \leq n^{-1/2} \log n \}$.

Theorem 2

Assume $p \in [1, 5/2)$, (A)–(E), (B*)–(D*). Let $l^\prime_{\theta \eta} = L^\prime_{\theta \eta}$. Then in $P_{\theta \eta}$-probability, the conditional distribution of
\[ n^{1/6} \left( n^{\beta/3} S'_{p_n} - \mu_p \int_0^1 |4\tilde{Z}_n'(t)\mu_{n}(t)|^{\beta/3} \, dt \right) \]

given \( X^{(n)} \) converges as \( n \to \infty \) to the Gaussian law with mean zero and variance given by (3).

As the Gaussian limit distribution is continuous, an immediate consequence of theorems 1 and 2 is that the bootstrap calibration is consistent.

**Corollary 2**
Assume \( p \in [1, 5/2) \), (A)–(E), (B*)–(D*). Let \( x \in (0, 1) \) and \( q_{p_n} \) be the \( x \)-upper percentile point of the conditional distribution of \( S'_{p_n} \) given \( X^{(n)} \). Then the test with critical region \( \{ S_{p_n} > q_{p_n}(x) \} \) has asymptotic level \( x \).

Consistency of the bootstrap deserves comments. Indeed, the naive bootstrap, which consists in drawing \( X_{n, 1}, \ldots, X_{n, n} \) for some \( q \in (0, 1) \), satisfies (A), and we consider an estimator \( \hat{\Lambda}_n \) as in (1). Typically, \( \hat{\theta}_n \) could be a maximum likelihood or a least-squares estimator under \( H_0 \). Note that for every \( \theta \) and \( \eta \) one then has

\[ \sup_{t \in [0, 1]} |\hat{\theta}_n(t) - \hat{\theta}_n(t)| = O_{P_{\theta, \eta}}(n^{-1/2}). \]  

We begin the section by briefly describing a slight modification of \( \hat{\Lambda}_n \) that is a more natural estimator than \( \hat{\Lambda}_n \) itself in some of the considered models.

4. Applications

In this section, we consider three specific models: in each model, we detail the estimator \( \Lambda_n \) on which the Grenander-type estimator \( \hat{\Lambda}_n \) is based, the construction of \( X^{(n)} \) and the assumptions under which the procedure is consistent. Hereafter, \( p \) denotes a fixed real in \([1, 5/2)\); we assume that the parametric model under the null hypothesis satisfies the smoothness assumptions (A) and (E) and we consider an estimator \( \hat{\theta}_n \) that satisfies (D). Typically, \( \hat{\theta}_n \) could be a maximum likelihood or a least-squares estimator under \( H_0 \).

4.1. A slight modification of the monotone estimator

Let \( \tilde{\Lambda}_n \) be the continuous version of \( \Lambda_n \), which means that \( \tilde{\Lambda}_n(t) = \Lambda_n(t) \) at every jump point \( t \) of \( \Lambda_n \) and at the boundaries \( t = 0 \) and \( t = 1 \), and \( \tilde{\Lambda}_n \) is linear in between two consecutive such points. The modification of the monotone estimator we consider is \( \tilde{\Lambda}_n \), the left-hand slope of the least concave majorant of \( \tilde{\Lambda}_n \), with \( \tilde{\Lambda}_n(0) = \lim_{t \to 0} \tilde{\Lambda}_n(t) \). Assume that for every \( \theta \) and \( \eta \),

\[ E_{\theta, \eta} \left( \sup_{t \in [0, 1]} |\tilde{\Lambda}_n(t) - \Lambda_n(t)|^q \right) \leq C n^{1-q} \]  

for some \( q > 12 \) and \( C > 0 \). Then, theorem 1 and corollary 1 remain true with \( \hat{\Lambda}_n \) replaced by \( \tilde{\Lambda}_n \). If, furthermore,

\[ E_{\theta} \left( \sup_{t \in [0, 1]} |\tilde{\Lambda}_n(t) - \Lambda_n(t)|^q \right) \leq C n^{1-q} \]  

for some \( q > 12 \) and \( C > 0 \).
with probability tending to one, where $\tilde{\Lambda}_n^*$ is computed in the same manner as $\tilde{\Lambda}_n$, but with $X^{(o)}$ replaced by $X^{(o)}$, then theorem 2 and corollary 2 remain true with $\hat{\lambda}_n$ replaced by $\lambda_n$. It can be shown that (9) and (10) hold in the models we consider next, so the following results remain true with $\hat{\lambda}_n$ replaced by $\lambda_n$.

4.2. Monotone density

Assume $X^{(o)} = (X_1, \ldots, X_n)$ where the $X_i$s are i.i.d. with a decreasing density $\lambda$ on [0, 1], and let $\Lambda_n$ be the corresponding empirical distribution function. Besides (A), (D) and (E), we assume:

(Ad) For every $\theta$, $\inf \lambda_\theta(t) > 0$.

Then, theorem 6 of Durot (2007) shows that (B) and (C) hold with $L_0 = \Lambda_0$ (as there is no nuisance parameter here, we omit the subscript $\eta$ in the notation). Setting $\hat{\tilde{\lambda}}_n(1) = 1$ and $\hat{\lambda}_n = \lambda_{\tilde{\lambda}_n}$, it follows from (8) and corollary 1 that the test with critical region (5) has asymptotic level $\alpha$.

For the bootstrap procedure, let $X^{(o)} = (X^*_1, \ldots, X^*_n)$, where the $X^*_i$s are conditional i.i.d. random variables with density function $\lambda_{\tilde{\lambda}_n}$. Assume, furthermore, that (D*) holds. It is shown in section I of the online Supporting Information that (B*) and (C*) hold with $L_n = \Lambda_{\tilde{\lambda}_n}$, so the test with critical region $\{S_{pn} > q_{\alpha,n}\}$ has asymptotic level $\alpha$.

4.3. Monotone regression

Assume one observes $X^{(o)} = (y_1, \ldots, y_n)$ with $y_i = \lambda(t_i) + \varepsilon_i$, $i = 1, \ldots, n$. Here, $t_i = i/n$, the $\varepsilon_i$s are i.i.d. random variables with mean zero and finite variance $\sigma^2$, $\lambda : [0, 1] \to \mathbb{R}$ is an unknown decreasing function, and the nuisance parameter is the common distribution $\eta$ of the $\varepsilon_i$s. Let $\Lambda_n$ be the partial sum process

$$\Lambda_n(t) = \frac{1}{n} \sum_{i \leq nt} y_i, \quad t \in [0, 1].$$

Besides (A), (D) and (E), we assume:

(Ar) There exists $q > 12$ such that $\mathbb{E}|\varepsilon_i|^q < \infty$, and $\mathbb{E}\varepsilon_i^2 = \sigma^2 > 0$.

The assumption that $q > 12$ here can probably be weakened at the price of more technicalities.

Let $\hat{\sigma}_n^2$ be an estimator of $\sigma^2$ such that $n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) = O_{P_\eta}(1)$ for all $\theta$ and $\eta$. One can consider for instance the Rice estimator

$$\hat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2;$$

(11)

other examples of suitable estimators can be found in Hall & Marron (1990). By theorem 5 of Durot (2007), (B) and (C) hold under our assumptions with $L_{\theta_0} = L_{\sigma^2}$. Setting $\hat{\tilde{\lambda}}_n(1) = \hat{\lambda}_n(t) = \hat{\sigma}_n^2$, it follows from corollary 1 that the test with critical region (5) has asymptotic level $\alpha$.

Now, we turn to the bootstrap calibration. Let $\hat{\lambda}_n$ be a non-parametric estimator of $\lambda$ that satisfies the following assumption.

(Ar*) For every $\theta$ and $\eta$, $\sup_t \left| \hat{\lambda}_n(t) \right| = O_{P_{\theta_0}}(1)$ and there exists an $s_0 > 3/4$ such that

$$\frac{1}{n} \sum_{i=1}^{n} \left( \hat{\lambda}_n(t_i) - \lambda_0(t_i) \right)^2 = O_{P_{\theta_0}}(n^{-2s_0/3}).$$

© 2010 Board of the Foundation of the Scandinavian Journal of Statistics.
One can consider for instance a kernel estimator. Alternatively, one can consider $\tilde{\lambda}_a := \tilde{\lambda}_a$ as it can be shown using theorem 2 of Durot (2007) that

$$n^{-1/3} \sum_{i=1}^{n} \left( \hat{\lambda}_a(t_i) - \lambda(t_i) \right)^2 = O_{\mathbb{P}_{\theta_0}}(1).$$

For every $i = 1, \ldots, n$, let

$$\tilde{e}_i = y_i - \tilde{\lambda}_a(t_i)$$

and

$$\tilde{e}_i = \tilde{e}_i - \frac{1}{n} \sum_{j=1}^{n} \tilde{e}_j.$$

Conditionally on $X^{(n)}$, let $e_1^*, \ldots, e_n^*$ be an $n$-sample from the distribution $\hat{\nu}_n$ that puts mass $1/n$ at each point $\tilde{e}_1, \ldots, \tilde{e}_n$. Then set $X^{(n)} = (y_1^*, \ldots, y_n^*)$, where for every $i = 1, \ldots, n$, $y_i^* = \tilde{\lambda}_a(t_i) + e_i^*$. We now have corollary 3.

**Corollary 3**

*In the regression setting where (A), (Ar), (D), (E), (D*) and (Ar*) hold, the test with critical region $\{S_{\nu_0} > q_{\nu_0}^{\alpha} \}$ has asymptotic level $\alpha$.***

**Proof.** Assume $\lambda = \lambda_0$ for some $\theta \in \Theta$. From (Ar) and (Ar*), there exists $C > 0$ such that

$$\mathbb{E}^* |e_i^*|^q = \frac{1}{n} \sum_{i=1}^{n} \tilde{e}_i - \frac{1}{n} \sum_{j=1}^{n} \tilde{e}_j |e_i^*|^q \leq C$$

with probability that tends to one. But there exists $C' > 0$ such that with probability that tends to one, $|\tilde{\lambda}_a(t)| \leq C'$ and $|\tilde{\lambda}_a(t)| \leq C'$ for all $t$, so

$$\sup_{t \in [0, 1]} |M_n^*(t) - \frac{1}{n} \sum_{i \leq m} e_i^*| \leq \frac{2C'}{n}$$

with probability that tends to one. Let $\sigma_n^{2*}$ be the common conditional variance of the variables $e_i^*$. Then, the assumption on $\tilde{\lambda}_a$ ensures that

$$\sigma_n^{2*} = \sigma^2 + O_{\mathbb{P}_{\theta_0}}(n^{-5/3}),$$

and one can prove that (B*) and (C*) hold with $L_n(t) = t\sigma_n^{2*}$ (see section II of the online Supporting Information for details). Therefore, corollary 3 follows from corollary 2.

**4.4. Right-censoring model with monotone hazard rate**

Assume one observes a right-censored sample $X^{(n)} = ((X_1, \delta_1), \ldots, (X_n, \delta_n))$. Here, $X_i = \min(T_i, Y_i)$ and $\delta_i = 1_{T_i \leq Y_i}$, where the $T_i$s are non-negative i.i.d. failure times and the $Y_i$s are i.i.d. censoring times independent of the $T_i$s. Let $\lambda$ be the restriction to $[0, 1]$ of the common hazard rate of the $T_i$s, that is, $\lambda = \mathbb{H}(1 - F)$ on $[0, 1]$, where $f$ and $F$ are the density and distribution function of $T_i$, respectively. Here, the nuisance parameter $\eta$ is the common distribution function of the $Y_i$s. Let $A_n$ be the restriction to $[0, 1]$ of the Nelson–Aalen estimator of $A$ and $\hat{\eta}_n$ be the Kaplan–Meier estimator of $\eta$. Besides (A), (D) and (E), we
make the following assumption, where \( F_0 \) denotes the distribution function of the law with hazard rate \( \lambda_0 \).

(Ac) \( \eta \) has a bounded continuous first derivative on \((0, 1)\) and \( \lim_{t \uparrow 1} \eta(t) < 1 \). Moreover, for every \( \theta \), \( \inf \lambda_\theta(t) > 0 \) and \( F_\theta(1) < 1 \).

By theorem 3 of Durot (2007), (B) and (C) hold with

\[
L_{\theta_0}(t) = \int_0^t \frac{\hat{\lambda}_\theta(u)}{(1 - F_\theta(u))(1 - \eta(u))} \, du.
\]

But \( \hat{\eta} \) uniformly converges to \( \eta \) in probability at the rate \( \sqrt{n} \), so with

\[
\hat{\lambda}_n = \frac{\hat{\lambda}_{\theta_0}}{(1 - F_{\theta_0})(1 - \hat{\eta}_n)} \quad \text{and} \quad \hat{\lambda}_n(1) = \int_0^1 \hat{\lambda}_n(t) \, dt,
\]

it follows from (8) and corollary 1 that the test with critical region (5) has asymptotic level \( \alpha \).

We now turn to the bootstrap calibration. Conditionally on \( X_1^{(n)}, \ldots, X_n^{(n)} \) be i.i.d. random variables with hazard rate \( \lambda_{\theta_0} \) and \( Y_1^{*}, \ldots, Y_n^{*} \) be i.i.d. random variables with distribution function \( \tilde{\eta}_n \), independent of the \( T_i \)'s. We define the bootstrap observation by

\[
X_i^{(n)} = (X_i^{(a)}, \delta_i), \quad (X_i^{(a)}, \delta_i) \sim (X_i^{(n)}, \delta_i),
\]

where \( X_i^{(n)} = \min(T_i, Y_i^{*}) \) and \( \delta_i = 1 \) if \( T_i < Y_i^{*} \). We assume, furthermore, that (D*) holds. Let \( \tilde{\eta}_n \) be the continuous version of \( \tilde{\eta}_n \) (we mean, the polygonal function that coincides with \( \tilde{\eta}_n \) at each jump point of \( \tilde{\eta}_n \) on \([0, 1]\) and such that \( \tilde{\eta}_n(1) = \lim_{t \downarrow 1} \tilde{\eta}_n(t) \)). Then,

\[
\sup_{t \in [0, 1]} |\tilde{\eta}_n(t) - \tilde{\eta}_n(t)| = O_{\Pr_{\theta_0}}(n^{-1}),
\]

so (see section III of the online Supporting Information for more details) similar arguments as before show that (B*) and (C*) hold with

\[
L^{*}_\theta(t) = \int_0^t \frac{\hat{\lambda}_{\theta_0}(u)}{(1 - F_{\theta_0}(u))(1 - \tilde{\eta}_n(u))} \, du.
\]

Therefore, the test with critical region \( \{S_{pn} > q_{\alpha/2}^n\} \) has asymptotic level \( \alpha \).

5. Simulation study

In this section, we report a simulation study performed in a regression model. Based on observations

\[
y_i = \lambda(t_i) + \sigma \varepsilon_i, \quad i = 1, \ldots, n, \quad (13)
\]

we study several tests for the simple null hypothesis

\[
H_0^\phi: \lambda = \lambda_0, \quad (14)
\]

where \( \lambda_0(t) = -10t, \ t \in [0, 1] \), and for the composite null hypothesis

\[
H_0^c: \lambda \in \{\lambda_\theta, \ \theta \in \Theta\}, \quad (15)
\]

where \( \Theta = \mathbb{R} \times (-\infty, 0) \) and for every \( \theta = (a, b) \), \( \lambda_\theta(t) = a + bt, \ t \in [0, 1] \). In (13), \( t_i = il/n \), the \( \varepsilon_i \)s are independently drawn from the standard Gaussian distribution, and we vary \( n, \sigma \) and \( \lambda \). The nominal level \( \alpha \) of the tests we investigate is of 5 per cent. These tests are described in section 5.1. Results about level and power are reported in sections 5.2 and 5.3, respectively.
5.1. Tests

Hereafter, \( \hat{\sigma}_n^2 \) denotes the Rice estimator of \( \sigma^2 \) given by (11), \( \hat{\theta}_n \) denotes the ordinary least-squares estimator of \( \theta \) given by

\[
\hat{b}_n = \frac{\sum_{i=1}^{n} (t_i - \hat{\theta})(y_i - \bar{y})}{\sum_{i=1}^{n} (t_i - \hat{\theta})^2} \quad \text{and} \quad \hat{a}_n = \bar{y} - \hat{b}_t,
\]

where \( \bar{t} \) and \( \bar{y} \) are the mean values of the \( t_s \) and \( y_s \), respectively, and \( \hat{\lambda}_n \) denotes the local linear kernel estimator (see Wand & Jones, 1995) with quartic kernel

\[
K(t) = \frac{15}{16} (1 - t^2)^2 1_{|t| \leq 1}
\]

and bandwidth denoted by \( h \). Every bootstrap calibration is performed using Monte Carlo simulations. Moreover, all the tests we investigate are presented next in the case of a composite null hypothesis \( H_0 \). The case of a simple null hypothesis \( H_0 \) is similar, replacing \( \hat{\lambda}_0 \) by \( \hat{\lambda}_0 \) in the formulae.

First, we fix \( p = 1 \) and consider the two tests described in section 4.3. We denote by \( \hat{S}_n \) and \( \hat{S}_n^* \) the tests based on plug-in and bootstrap calibrations, respectively, where \( 8k_1 \) and \( 2\mu_1 \) are approximated by 0.17 and 0.82, respectively, and the bootstrap calibration is performed using the local linear kernel estimator \( \hat{\lambda}_n \) with a bandwidth \( h \) selected by cross-validation. Note that for time cost necessity, we do not minimize the cross-validation function over the whole interval \([0, 1]\) but over an interval \( I \subset [0, 1] \) suggested by a preliminary study. Moreover, to assess the qualities of cross-validation in our setting, we also implement the bootstrap calibration where \( h \) is fixed to a value \( h_0 \) depending on the unknown function \( \lambda \), and suggested by a preliminary study of the estimated mean-squared error. We denote by \( \hat{S}_n^{*m} \) the corresponding test (which, of course, cannot be implemented in practice as it relies on the knowledge of \( \lambda \)).

Next, we consider three tests that are similar to \( \hat{S}_n^* \), but that are calibrated using different bootstrap procedures. Their study attempts to justify the choice of our bootstrap calibration method. The first two modifications consist in replacing the local linear kernel estimator \( \hat{\lambda}_n \) in (12) by either the parametric estimator \( \hat{\lambda}_0 \) or the Grenander-type estimator \( \hat{\lambda}_n \), so that the resulting test does not depend on a smoothing parameter. Apart from this replacement, the tests are identical to \( \hat{S}_n^* \). The first proposal is denoted by \( \hat{S}_n^{*m} \) in the sequel, and the second one is denoted by \( \hat{S}_n^{*m} \). The third modification we consider consists in studentization: it rejects the null hypothesis when

\[
\frac{1}{\hat{\sigma}_n \sqrt{0.17}} \left( n^{1/3} S_{1n} - 0.82 \hat{\sigma}_n^{2/3} \int_0^1 |\hat{\lambda}_{\hat{\theta}}(t)/2|^{1/3} dt \right)
\]

exceeds the \( \alpha \)-upper percentile point of its bootstrap version, obtained by replacing \( S_{1n} \), \( \hat{\sigma}_n \) and \( \hat{\theta}_n \) by the same quantities based on the bootstrap observation (which is generated as in section 4.3 with \( \hat{\lambda}_n \) as the local linear kernel estimator). This test is denoted by \( Z_n \) in the sequel.

Finally, we consider three competitor tests. Two of them are goodness-of-fit tests that do not take into account the monotonicity constraint. The first one, which is denoted by \( T_n^{*o} \) in the sequel, is similar to ours, but with the Grenander-type estimator \( \hat{\lambda}_n \) replaced by the local linear kernel estimator \( \hat{\lambda}_n \): it consists in rejecting the null hypothesis when

\[
\int_0^1 (\hat{\lambda}_n(t) - \hat{\lambda}_{\hat{\theta}}(t))^2 dt
\]
exceeds the $z$-upper percentile point of its bootstrap version, obtained by replacing $\tilde{\lambda}_n(t)$ and $\hat{\theta}_n$ by the same quantities based on the bootstrap observation, generated as in section 4.3. For time cost reasons, we do not select the bandwidth $h$ using cross-validation. Rather, similar to $S_{n^m}$, $h$ is fixed here to a value $h_0$ depending on the unknown function $\lambda$, suggested by a preliminary study of the estimated mean-squared error. The second competitor is the Kolmogorov–Smirnov test studied by Stute et al. (1998) and is denoted by $D_n^*$ in the sequel. The last competitor test is the smooth test proposed by Ducharme & Fontez (2004). It is denoted by $R_n$ in the sequel. To our knowledge, it is the only goodness-of-fit test specifically designed for monotone regression functions available in the literature. However, $R_n$ does not allow testing $H_{s0}$. Indeed, it is designed to test a parametric model for $f := \lambda'(\lambda(1) - \lambda(0))$, where $\lambda'$ is the derivative of $\lambda$. Thus, it allows testing $H_{c0}$, which is equivalent to $f = 1$, but does not allow testing $H_{s0}$. Moreover, it requires the regression function to be increasing and positive, so we change $y_i$ into $-y_i$ to implement the method. The test statistic is defined by (13) in Ducharme & Fontez (2004). The parameter $K$ in this formula is optimized as indicated, taking $d = 1$ and $D = 8$, and the critical value is given in table 1 of that paper.

5.2. Levels

Hereafter, we investigate the empirical levels of the aforementioned tests. For each pair $(n, \sigma)$, $n \in \{50, 100\}$, $\sigma \in \{0.5, 1\}$, we generate observations (13) with $\lambda = \tilde{\lambda}_0$. Then for each test, we compute the test statistic and the critical value. In the case of a bootstrap calibration, the critical value is computed using 1000 bootstrap replications. The empirical levels of the tests are then defined as the percentage of rejection of the null hypothesis over 10,000 replications. They are reported in Table 1 for $\sigma = 0.5$. For the sake of briefness, the levels obtained for $\sigma = 1$ are not displayed here, but they lead to similar conclusions.

It can be seen in Table 1 that $\hat{\lambda}_n$ is very anticorrelative. This means that for moderate sample sizes, the asymptotic normal approximation is misleading (as already noticed by Durot and Tocquet in the simple null hypothesis case). Similarly, $S_{n^m}$ is misleading. The other tests show empirical levels quite close to the nominal level of 5 per cent. However, $S_{n^p}$ tends to be conservative when testing $H_{c0}$, and $S_{n^m}$, which behaves similarly as $S_{n^p}$, is not very accurate for small sample sizes ($n = 50$). This drawback is overcome by studentization: the empirical level of $Z_n^*$ is very close to 5 per cent in every considered case. Both $D_n^*$ and $R_n$ perform well for testing $H_{c0}$, and $D_n^*$ is slightly anticonservative when testing $H_{s0}$. The same drawback is observed for $T_n^*$, for both simple and composite hypotheses.

Table 1. Empirical levels (in per cent) of the tests for $H_{s0}^*$ and $H_{c0}^*$ when $\sigma = 0.5$

<table>
<thead>
<tr>
<th>Test</th>
<th>$H_{c0}^*$</th>
<th>$H_{s0}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=50$</td>
<td>$n=100$</td>
<td>$n=50$</td>
</tr>
<tr>
<td>$\hat{\lambda}_n$</td>
<td>17.7</td>
<td>15.5</td>
</tr>
<tr>
<td>$S_{n^m}$</td>
<td>6.5</td>
<td>11.5</td>
</tr>
<tr>
<td>$S_{n^p}$</td>
<td>4.5</td>
<td>4.6</td>
</tr>
<tr>
<td>$S_n^*$</td>
<td>5.8</td>
<td>5.2</td>
</tr>
<tr>
<td>$S_{n^m}$</td>
<td>5.7</td>
<td>5.2</td>
</tr>
<tr>
<td>$Z_n^*$</td>
<td>5.1</td>
<td>5.2</td>
</tr>
<tr>
<td>$D_n^*$</td>
<td>5.9</td>
<td>5.7</td>
</tr>
<tr>
<td>$T_n^*$</td>
<td>7.0</td>
<td>6.2</td>
</tr>
<tr>
<td>$R_n$</td>
<td>4.9</td>
<td>4.8</td>
</tr>
</tbody>
</table>

© 2010 Board of the Foundation of the Scandinavian Journal of Statistics.
5.3. Powers

We now investigate the empirical powers of the tests. As the empirical levels of \( \hat{S}_n \) and \( S_n^{\alpha} \) are much larger than the nominal level of 5 per cent, we did not study the power of these tests and concentrate on the other ones. We focus on alternative functions of the form

\[
\lambda(t) = \lambda_0(t) + c\Delta(t),
\]

where

\[
\Delta(t) = \delta^{-112} \left( 0.25 - \left( \frac{t - t_0}{\delta} - 0.5 \right) \right)^3 \mathbb{1}_{t_0 \leq t \leq t_0 + \delta}.
\]

Here, \( c, \delta \) and \( t_0 \) are positive constants. The monotonicity constraint on \( \lambda \) implies imposing \( c < \frac{1}{32} \sqrt{\frac{1}{2}} \). The functions (16) are obtained from the straight line \( \lambda_0 \) by an additional perturbation which takes the form of a local bump. The constants \( c \) and \( \delta \), respectively, adjust the height and width of the bump and the value of \( t_0 \) is chosen such that the bump is centred. We consider here the cases \( (\delta, t_0) = (0.3, 0.35) \) and \( c \in \{0.2, 0.35\} \) and \( c \in \{5, 10, 15\} \) for each pair \( (n, \sigma) \), \( n \in \{50, 100\} \), \( \sigma \in \{0.5, 1\} \) and each considered alternative function, we generate observations (13), with \( \lambda \) defined by (16). Then for each test, we compute the test statistic and the critical value. The critical value of each test relying on a bootstrap calibration is computed using 800 bootstrap replications. The empirical powers of the tests are then defined as the percentage of rejection over 8000 replications. Performances are reported in Tables 2 and 3 for \( H^o_0 \) and \( H^c_0 \), respectively. We only display the results for \( \sigma = 0.5 \), but the following comments rely on the whole study \( (\sigma = 0.5 \) and \( \sigma = 1) \).

For each pair \( (n, \sigma) \), \( n \in \{50, 100\} \), \( \sigma \in \{0.5, 1\} \) and each considered alternative function, we generate observations (13), with \( \lambda \) defined by (16). Then for each test, we compute the test statistic and the critical value. The critical value of each test relying on a bootstrap calibration is computed using 800 bootstrap replications. The empirical powers of the tests are then defined as the percentage of rejection over 8000 replications. Performances are reported in Tables 2 and 3 for \( H^o_0 \) and \( H^c_0 \), respectively. We only display the results for \( \sigma = 0.5 \), but the following comments rely on the whole study \( (\sigma = 0.5 \) and \( \sigma = 1) \).

It is seen in Tables 2 and 3 that the power of the tests increases with the sample size and the height \( c \) of the bump. Over the whole set of results, it appears that \( S^p_n \) and \( S^{\alpha}_n \) behave similarly. They dominate \( S^p_n \) against all the considered alternatives, and also \( Z_n \) against most of the considered alternatives: \( Z_n \) is slightly more powerful against small departures from a composite null hypothesis but is outperformed by \( S^p_n \) and \( S^{\alpha}_n \) in the other cases. This justifies the choice of our bootstrap method and the use of cross-validation. Globally, \( R_n \) is the less powerful test, and \( D^*_n \) is often outperformed by either \( T^{\alpha}_n \) or \( S^*_n \) (or even by both these tests). Actually, \( S^*_n \) (and \( S^{\alpha}_n \)) appears to dominate every other test in the case of a simple

<table>
<thead>
<tr>
<th>( (\delta, t_0) )</th>
<th>Test</th>
<th>( n = 50 )</th>
<th>( c = 10 )</th>
<th>( c = 20 )</th>
<th>( c = 30 )</th>
<th>( n = 100 )</th>
<th>( c = 10 )</th>
<th>( c = 20 )</th>
<th>( c = 30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (0.3, 0.35) )</td>
<td>( S^p_n )</td>
<td>11.8</td>
<td>27.4</td>
<td>49.1</td>
<td>17.0</td>
<td>51.1</td>
<td>85.1</td>
<td>10.2</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td>( S^*_n )</td>
<td>14.6</td>
<td>52.1</td>
<td>81.4</td>
<td>19.0</td>
<td>70.4</td>
<td>95.8</td>
<td>12.2</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>( S^{\alpha}_n )</td>
<td>15.6</td>
<td>53.0</td>
<td>81.2</td>
<td>20.4</td>
<td>68.6</td>
<td>96.5</td>
<td>12.2</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>( Z_n )</td>
<td>12.5</td>
<td>53.0</td>
<td>81.2</td>
<td>20.4</td>
<td>68.6</td>
<td>96.5</td>
<td>12.5</td>
<td>53.0</td>
</tr>
<tr>
<td></td>
<td>( D_n^* )</td>
<td>10.2</td>
<td>20.4</td>
<td>36.9</td>
<td>13.3</td>
<td>37.2</td>
<td>66.5</td>
<td>8.4</td>
<td>45.3</td>
</tr>
<tr>
<td></td>
<td>( T^{\alpha}_n^* )</td>
<td>6.2</td>
<td>10.0</td>
<td>15.4</td>
<td>7.5</td>
<td>12.9</td>
<td>25.8</td>
<td>7.9</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>( S^{\alpha}_n )</td>
<td>6.7</td>
<td>10.7</td>
<td>17.8</td>
<td>8.0</td>
<td>14.1</td>
<td>28.5</td>
<td>5.9</td>
<td>8.3</td>
</tr>
</tbody>
</table>
null hypothesis, whereas $T^{*o}$ tends to outperform the others in the composite null hypothesis case, except for some cases of small departures, where $D^*$ achieves a slightly better power.

6. Conclusion

In this article, we have presented a test of a composite null hypothesis which is specifically designed for monotone functions. The test statistic is based on a non-parametric data-driven estimator of the function under study, so that no extra parameter has to be adjusted to compute the test statistic. To calibrate our test, we have proposed a general bootstrap method covering three statistical frameworks: regression, density and hazard rate functions. We believe that similar constructions could be applied to deal with other statistical models.

Acknowledgements

The authors are grateful to the editor and the referees for their insightful comments to improve the quality of the article.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Details on the results in section 4 are available as supporting information for this article. This material may be found in the online version of the article.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References


© 2010 Board of the Foundation of the Scandinavian Journal of Statistics.


Received January 2008, in final form December 2009

Cécile Durot, Université Paris-Sud, Bâtiment 425, 91405 Orsay Cedex, France. E-mail: cecile.durot@math.u-psud.fr

Appendix

This appendix is devoted to the proof of theorem 2. The similar proof of theorem 1 is omitted. Note that theorem 2 is a generalization of theorem 2 in Durot (2007) as it provides
the asymptotic distribution of the $\mathbb{L}_p$-distance between a monotone estimator and a monotone function. The generalization is twofold. First, the considered monotone estimator $\hat{\lambda}_n$ is computed from an observation $X^{(\theta)}$ the distribution of which depends on parameters $\lambda_{\delta_n}$ and $\hat{\eta}_n$ that are allowed to depend on $n$. As a consequence, the function $L_n$ that appears in the assumption (B*) may also depend on $n$. Second, we do not consider here the distance between the monotone estimator and the true underlying function $\hat{\lambda}_{\delta_n}$ itself, but we consider the distance between the monotone estimator and an estimator of the true underlying function. Despite these differences, the proof of theorem 2 follows the lines of the proof of Durot’s theorem 2. Therefore, details are omitted and we only stress the places where the two proofs differ. Moreover for notational convenience, we sometimes omit subscript $n$ and also subscripts $\theta$ and $\delta$. Thus, in the sequel, we assume $\lambda = \lambda_{\theta}$ for some $\theta$ and we write $L$ and $l$ for $L_{\theta}$ and $l_{\theta}$, respectively.

Under assumptions (A), (D) and (E), there exist positive reals $c$ and $C$ that do not depend on $n$ such that the probability that

$$\sup_{t \in [0, 1]} |\lambda_g(t)| \leq C, \quad \inf_{t \in [0, 1]} |\lambda_g(t)| > c, \quad \sup_{t \in [0, 1]} |\lambda'_g(t)| \leq C$$

(17)

tend to one as $n \to \infty$. But to prove theorem 2, we can restrict ourselves to an event whose probability tends to one, so we can assume without loss of generality that (17) holds. Likewise, thanks to assumptions (7) and (D), we assume

$$\sup_{t \in (0, 1)} |l(t) - l'(t)| \leq Cn^{-\alpha/3} \log n \quad \text{and} \quad \| \hat{\theta} - \theta \| \leq n^{-1/2} \log n.$$  

- **Step 1.** For every $a \in \mathbb{R}$, let

$$\hat{U}^*(a) = \text{argmax}_{a \in [0, 1]} \{ L_n^*(a) - au \},$$

(18)

where argmax denotes the greatest location of maximum. Let $g_\hat{\theta}$ and $g_\theta^*$ be the inverse functions of $\hat{\lambda}_{\hat{\theta}}$ and $\lambda_{\theta^*}$, respectively. Arguing as for Durot’s lemmas 3 and 4 one obtains that there exists $K > 0$ such that

$$\mathbb{P}^* [ |\hat{U}^*(a) - g_\theta(a) | \geq x ] \leq K (nx^3)^{1-\alpha}$$

(19)

for every $a \in \mathbb{R}$ and $x > 0$, and

$$\mathbb{P}^* [ |\hat{U}^*(a) - g_\theta(a) | \geq x ] \leq \frac{K}{nx (\hat{\lambda}_g(g_\theta(a)) - a)^2}$$

(20)

for every $x > 0$ and $a \notin \lambda_{\theta}([0, 1])$. Arguing as for Durot’s theorem 1 one then obtains that for every $p' \in [1, 2)$, there exists $K > 0$ such that

$$\mathbb{E}^* [ \hat{\lambda}_g(t) - \lambda_{\theta}(t) ]^{p'} \leq K n^{-p'/3}$$

for all $t \in [n^{-1/3}, 1 - n^{-1/3}]$ and

$$\mathbb{E}^* [ \hat{\lambda}_g(t) - \lambda_{\theta}(t) ]^{p'} \leq K [n(t \land (1 - t))]^{-p'/2}$$

for all $t \in (0, n^{-1/3}] \cup [1 - n^{-1/3}, 1)$. Now, set $x_+ = x \vee 0$ for every $x \in \mathbb{R}$ and let

$$I_1 = \int_0^1 \left( \hat{\lambda}_g(t) - \lambda_{\theta}(t) ight)_+^p \, dt = \int_0^1 \int_0^{\infty} 1_{\hat{\lambda}_g(t) > \lambda_{\theta}(t) + x^p} \, da \, dt.$$

We have $\hat{\lambda}_g(t) < \lambda_{\theta}^*(0)$ for all $t > \hat{U}^*(\lambda_{\theta}^*(0))$; so
Therefore, \( I_{n}^{*} = \int_{0}^{1} \int_{0}^{L_{n}^{*}(\theta(r))^{q}} I_{\theta}(t) \geq \theta_{0}(t) + a_{0}^{1/p} \, dt + R_{n}^{*} \),

where

\[
0 \leq R_{n}^{*} \lesssim \int_{0}^{L_{\theta}(\theta_{0}(0))} \left( \frac{\hat{L}^{*}(t) - \hat{\theta}(t)}{t} \right)^{q} \, dt.
\]

But \( \hat{U}^{*} \) is non-increasing and \( \hat{\theta}_{0}(0) = \lambda_{0}(0) + O_{p^{*}}(n^{-1/2}) \); so, we have

\[
P^{*} \left( \hat{U}^{*}(\lambda_{0}(0)) > n^{-1/2} \log n \right) \leq P^{*} \left( \hat{U}^{*}(\lambda_{0}(0) - n^{-1/2} \log n) > n^{-1/2} \log n \right) + o(1).
\]

Similar to step 1 in Dutro (2007), (19) then implies

\[
R_{n}^{*} \leq \int_{0}^{n^{-1/2} \log n} \left( \hat{L}^{*}(t) - \hat{\theta}(t) \right)^{q} \, dt + o_{p^{*}}(n^{-p/3 - 1/6}),
\]

and as \( \hat{\theta} = \hat{\theta}_{0} + O_{p^{*}}(n^{-1/2}) \),

\[
R_{n}^{*} = o_{p^{*}}(n^{-p/3 - 1/6}).
\]

Change of variable \( b = \hat{\theta}_{0}(t) + a_{0}^{1/p} \) then yields

\[
I_{n}^{*} = \int_{\hat{\theta}_{0}(0)}^{\hat{\theta}_{0}(1)} \int_{g_{0}^{*}(b)}^{g_{0}^{*}(b - t)} p(b - \hat{\theta}_{0}(t))^{q-1} \, dt \, db + o_{p^{*}}(n^{-p/3 - 1/6}).
\]

It can be shown using Taylor’s expansion that there exists \( K > 0 \) such that for every \( b \) in the range of \( \hat{\theta}_{0}^{*} \),

\[
|b - \hat{\theta}_{0}^{*}(t)|^{q-1} - |(g_{0}^{*}(b) - t)\hat{\theta}_{0}^{*}(b)|^{q-1} | \leq K|t - g_{0}^{*}(b)|^{q-1} (o_{p^{*}}(n^{-1/6}) + |t - g_{0}^{*}(b)|^{q})
\]

where \( o_{p^{*}} \) is uniform in \( t \) and \( b \). Moreover, integrating (19) proves that there exists \( K > 0 \) such that for every \( \varphi' < 3(q - 1) \) and \( a \in \mathbb{R} \),

\[
\mathbb{E}^{*} \left( n^{1/3} |\hat{U}^{*}(a) - g_{0}^{*}(a)|^{q} \right) \leq K.
\]

Therefore, \( I_{n}^{*} \) is asymptotically equivalent to

\[
\int_{\hat{\theta}_{0}(0)}^{\hat{\theta}_{0}(1)} \int_{g_{0}^{*}(b)}^{g_{0}^{*}(b - t)} p(t - g_{0}^{*}(b))^{q-1} |\hat{\theta}_{0}(b)|^{q-1} \, dt \, db.
\]

Hence, \( I_{n}^{*} \),

\[
S_{nm}^{*} \leq \int_{\lambda_{0}(0)}^{\lambda_{0}(1)} \left| \hat{U}^{*}(a) - g_{0}^{*}(a) \right|^{p} \left| g_{0}^{*}(a) \right|^{1-p} \, da + o_{p^{*}}(n^{-p/3 - 1/6}).
\]

• Step 2. We have

\[
\sup_{a \in (\lambda_{0}(0), \lambda_{0}(1))} \left| g_{0}^{*}(a) - g_{0}^{*}(a) \right| = O_{p^{*}}(n^{-1/2}),
\]

so by (22), \( S_{nm}^{*} \) is asymptotically equivalent to

\[
\int_{\hat{\theta}_{0}(0)}^{\hat{\theta}_{0}(1)} \left| L^{*}(\hat{U}(a)) - L^{*}(g_{0}^{*}(a)) + (g_{0}^{*}(a) - g_{0}^{*}(a)) l (g_{0}^{*}(a)) \right|^{p} \left| g_{0}^{*}(a) \right|^{1-p} \, da.
\]
We have the representation $B^*_n(t) = W^*(t) - \xi^* t$, where $W^*$ is a standard Brownian motion, $\xi^* \equiv 0$ if $B^*_n$ is a Brownian motion and $\xi^*$ is a standard Gaussian variable independent of $B^*_n$ if $B^*_n$ is a Brownian bridge. For every $a \in \mathbb{R}$, let

$$a^* = a - n^{-\frac{1}{2}}(g_{\beta}(a)).$$

Similar to Durot’s formula (27), one gets

$$\mathbb{P}^x(\{L^* (\tilde{U}^*(a^*)) - L^* (g_{\beta}(a))\} > x) \leq K(nx^{1-q})$$

(23)

for some $K > 0$ and all $x > 0$. Integrating this inequality proves that

$$\sup_{a \in \mathbb{R}} E \left[ \left( n^{1/3} |L^* (\tilde{U}^*(a^*)) - L^* (g_{\beta}(a))| \right)^q \right] \leq K$$

for some $K > 0$ provided $q' < 3(q-1)$. Using again similar arguments as in Durot’s step 2, one can then perform the change of variable $a \to a^*$ to get that $S_{pm}^*$ is asymptotically equivalent to

$$\int_{\tilde{U}^*(a^*)} W_{g_{\beta}(a)}(u) + R^*(a, u)$$

(24)

over

$$I^*(a) = \left[ n^{1/3}(L^*(0) - L^*(g_{\beta}(a))), n^{1/3}(L^*(1) - L^*(g_{\beta}(a))) \right],$$

where for every $a$ and $u$,

$$D^*(a, u) = n^{2/3} \left( \Lambda_{\beta} \circ L^{*-1} - aL^{*-1} \right) (L^*(g_{\beta}(a))) + n^{-1/3} u$$

$$- n^{2/3} (\Lambda_{\beta}(g_{\beta}(a)) - g_{\beta}(a)),$$

$$W_{g_{\beta}(a)}(u) = n^{1/6} \left[ W^* \left( L^*(g_{\beta}(a)) + n^{-1/3} u \right) - W^* \left( L^*(g_{\beta}(a)) \right) \right]$$

and where with probability that tends to one, the remainder term $R^*(a, u)$ satisfies

$$|R^*(a, u)| \leq n^{2/3} \sup_{t \in [0, 1]} |\Lambda^*_n(t) - \Lambda_{\beta}(t) - n^{-1/2} B^*_n(t)|$$

$$+ Kn^{-1/6} |\xi^* u| (Cn^{-2/3} \log n + n^{-1/3} |u|).$$

Let

$$\tilde{U}^*(a) = \arg\max_{u \in [-\log n, \log n]} \{ D^*(a, u) + W_{g_{\beta}(a)}(u) \}.$$

Similar arguments as in step 2 of Durot (2007) then show that

$$n^{2/3} S_{pm}^* = \int_{\tilde{U}^*(a^*)} \left( \tilde{U}^*(a) + n^{1/3} (g_{\beta}(a) - g_{\beta}(a^*)) l(g_{\beta}(a)) \right) \left| \frac{g_{\beta}'(a)}{l(g_{\beta}(a))} \right|^{1-p} \, da$$

$$+ o_p(n^{-1/6}).$$

Uniformly in $a$ and $u \in [-\log n, \log n]$, $D^*(a, u)$ can be approximated by
\[ n^{2/3} \left( \Lambda_\delta \circ L^{-1} - aL^{-1} \right) \left( L(g_{\delta}(a)) + n^{-1/3}u \right) - n^{2/3} \left( \Lambda_\delta(g_{\delta}(a)) - ag_{\delta}(a) \right) \]

which, in turn, can be approximated by

\[ \frac{\hat{x}'(g_{\delta}(a))}{2(l(g_{\delta}(a)))^{2/3}} u^2, \]

the remainder being of the order \( O(n^{-4/3/\log n}) \). Defining

\[ \hat{V}^*(t) = \arg\max_{-\log n < \log n} \left\{ -\frac{\hat{x}'(t)}{2l(t)} u^2 + W^*_n(u) \right\} \]

and using similar arguments as in Durot (2007), one then gets

\[ p^{3/2} S_{3n} = \int_{\hat{x}'(0)}^{\hat{x}'(1)} \left| \hat{V}^*(g_{\delta}(a)) + n^{1/3} \left( g_{\delta}(a) - g_{\delta}(a') \right) l(g_{\delta}(a)) \right|^p \frac{|\hat{x}'(a)|^{1-p}}{|l(g_{\delta}(a))|} da \]

\[ + o_{p^*}(n^{-1/6}) \]

\[ = \int_0^1 \left| \hat{V}^*(t) + n^{1/3} \frac{l(t)}{\hat{x}'(t)} \left( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \right|^p \frac{\hat{x}'(t)}{l(t)} dt \]

\[ + o_{p^*}(n^{-1/6}). \]

- **Step 3.** Let \( H = |\hat{x}'|^{1/p} \) and \( D^*_n \) be equal to

\[ \int_0^1 \left( \left| \hat{V}^*(t) + n^{1/3} \frac{l(t)}{\hat{x}'(t)} \right( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \right|^p \frac{\hat{x}'(t)}{l(t)} dt \]

We show here that

\[ D^*_n = o_{p^*}(n^{-1/6}). \]  

(25)

One can approximate \( D^*_n \) by

\[ \int_0^1 \left( \left| \hat{V}^*(t) + n^{1/3} \frac{l(t)}{\hat{x}'(t)} \right( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \right|^p \frac{\hat{x}'(t)}{l(t)} dt \]

as the distance between \( D^*_n \) and this integral is less than

\[ \sup_{\|\delta - \hat{\theta}\| \leq \varepsilon} \sup_{\|\hat{\theta} - \theta\| \leq \varepsilon} \|\hat{x}_0(t) - \hat{\lambda}_0(t)\| \cdot O_{p^*}(n^{-1/6}) = o_{p^*}(n^{-1/6}). \]

We have

\[ \sup_{\varepsilon} \left| n^{1/3} l(t) \left( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \right| = O_{p^*}(n^{-1/6}) \]

and \( \hat{V}^*(t) \) possesses uniformly bounded moments of any order so, similar to Durot (2007), expanding \( x \mapsto x^p \) around \( |\hat{V}^*(t)| \) proves that \( D^*_n \) is asymptotically equivalent to

\[ p \int_0^1 \left( \left| \hat{V}^*(t) + n^{1/3} \frac{l(t)}{\hat{x}'(t)} \right( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \right)^p \frac{\hat{x}'(t)}{l(t)} dt \]

For every \( x \geq 0 \) and \( y > 0 \) we have \( x - y = (x^2 - y^2)/(x + y) \), so

\[ D^*_n = p \int_0^1 2n^{1/3} l(t) \left( \hat{x}_0(t) - \hat{\lambda}_0(t) + n^{-1/2} \hat{x}_0'' \right) \frac{\hat{V}^*(t)}{2\hat{x}'(t)\hat{V}^*(t)} \left| \hat{V}^*(t) \right|^{p-1} H(t) dt \]

\[ + o_{p^*}(n^{-1/6}), \]
where $O_{p^*}(n^{-1/6})$ is uniform in $t$. Hence,

$$D_n^* = p \int_0^1 n^{1/3} \frac{l(t)}{\bar{x}(t)} \left( \hat{\beta}(t) - \hat{\theta} \right) \bar{V}^*(t) \left| \bar{V}^*(t) \right|^{p-2} H(t) \, dt + o_{p^*}(n^{-1/6}).$$

Similar to step 3 in Durot (2007) one obtains that for every non-random function $f_n$ such that $\sup_n |f_n(t)|$ is bounded,

$$\int_0^1 \bar{V}^*(t) \left| \bar{V}^*(t) \right|^{p-2} f_n(t) \, dt = o_{p^*}(1).$$

Therefore,

$$D_n^* = (\hat{\theta} - \hat{\theta}) o_{p^*}(n^{1/3}) + \tilde{\xi} o_{p^*}(n^{-1/6}),$$

which yields (25).

- **Step 4.** From the preceding steps,

$$np^{1 \over 3} S_m^* = \int_0^1 \left| \bar{V}^*(t) \right|^{p} \left| \frac{\dot{\beta}(t)}{l(t)} \right|^{p} \, dt + o_{p^*}(n^{-1/6}).$$

One then obtains the required result by using the same arguments as in steps 4, 5 and 6 of Durot (2007).