Depth estimators and tests based on the likelihood principle with application to regression*

by Christine H. Müller
Carl von Ossietzky University of Oldenburg

April 15, 2004

Abstract

We investigate depth notions for general models which are derived via the likelihood principle. We show that the so-called likelihood depth for regression in generalized linear models coincides with the regression depth of Rousseeuw and Hubert (1999) if the dependent observations are appropriately transformed. For deriving tests, the likelihood depth is extended to simplicial likelihood depth. The simplicial likelihood depth is always a U-statistic which is in some cases not degenerated. Since the U-statistic is degenerated in the most cases, we demonstrate that nevertheless the asymptotic distribution of the simplicial likelihood depth and thus asymptotic α-level tests for general types of hypotheses can be derived. The tests are distribution-free. We work out the method for linear regression with and without intercept and for quadratic regression.

Keywords: Likelihood depth, simplicial depth, regression depth, generalized linear models, logistic regression, Poisson distribution, exponential distribution, polynomial regression, degenerated U-statistic, distribution-free tests, spectral decomposition.
AMS Subject classification: Primary 62G05, 62G10; secondary 62J05, 62J12, 62G20.

*Research supported by the grant Mu 1031/6-1 of the Deutsche Forschungsgemeinschaft.

1 Introduction

For generalizing the median to multivariate data sets, maximum depth estimators based on different depth notions have been introduced. Different depth notions are, for example, the half space depth of Tukey (1975) and the simplicial depth of Liu (1988, 1990). For other depth notions see the book of Mosler (2002) and the references in it. Multivariate depth concepts were transferred to regression by Rousseeuw and Hubert (1999), to logistic regression by Christmann and Rousseeuw (2001) and to the Michaelis-Menten model by Van Aelst et al. (2002).

Since many depth concepts exist, there are attempts to provide a general theory for them. Zuo and Serfling (2000a) proposed properties which are desirable for depth notions. In Zuo and Serfling (2000b), it is shown that these desirable properties ensure well behaved contours and almost sure convergence. While Zuo and Serfling provided a general theory via some properties, Mizera (2002) introduced a general definition of depth by using general objective (criterial) functions and constructed a differential approach for it. Especially, the half space depth of Tukey (1975) and the regression depth of Rousseeuw and Hubert (1999) are special cases of the general definition. Although the approach of Mizera (2002) holds for general objective functions, the objective functions, given in that paper by examples, all base on residuals, i.e. on $y_n - \theta$ or $y_n - x_n^\top \theta$. But they also can be based on likelihood functions as Mizera and Müller (2003) pointed out. They worked out this possibility for simultaneous estimation of location and scale leading to location-scale depth.

In this paper in Section 2, the approach of likelihood depth, where the objective function and thus the depth notion is based on the likelihood function, is studied for a broader class of applications. Likelihood depth is worked out for regression in generalized linear models as logistic regression and regression with Poisson distribution, geometric distribution and exponential distribution. It is shown that in all cases the depth notion is equivalent to the regression depth of Rousseeuw and Hubert (1999) if the dependent observations are transformed appropriately. This means that the depth in these generalized linear models has the same robustness properties as the regression depth of Rousseeuw and Hubert (1999) and can be calculated like this.

In Section 2 it is also shown that the half space depth of Tukey (1975) is a likelihood depth. Since the simplicial depth of Liu (1988, 1990) is an extension of the half space depth we also define simplicial likelihood depth as extension of the likelihood depth in this section. We are aware of the fact that simplicial depth and thus also simplicial likelihood depth possesses not all of the desirable properties proposed by Zuo and Serfling (2000a). But simplicial depth and thus simplicial likelihood depth has the strong advantage that the depth function is a U-statistic. For U-statistics, the asymptotic distribution can be derived rather easily. Unfortunately, the simplicial depth for multivariate location is a degenerated U-statistic as Liu (1990) pointed out. Hence the asymptotic distribution is not that easy to derive.

Arcones et al. (1994) derived the asymptotic normality of the maximum simplicial depth estimator via the convergence of the whole U-process. The convergence of the U-process was also shown by Dümbgen (1992). However the asymptotic normal distribution has a covariance matrix which depends on the underlying distribution. Hence this result cannot be used to derive distribution-free tests, a hope which is related to the introduction of depth notions since the depth generalizes the rank of a one-dimensional observation. Therefore Liu (1992), Liu and Singh (1993) proposed a different approach for deriving distribution-free multivariate rank tests based on depth notions. It generalizes Wilcoxon's rank sum test for two samples. While the asymptotic normality is derived for several depth notions for distributions on $I\!R^1$, it is shown only for the Mahalanobis depth for distributions on $I\!R^k$, $k > 1$. Hence it is unclear how to generalize the approach of Liu and Singh to other situations as regression. Tests for regression based on depth notions are derived only by Van Aelst et al. (2002). They even derived an exact test based on the regression depth but did it only for linear regression.

In this paper in Section 3 we derive simple distribution-free tests for regression based on the simplicial likelihood depth. These tests can test all hypothesis of the form $H_0 : \theta \in \Theta_0$ where Θ_0 is a subset of the parameter space and are not restricted to regression problems. It is a general approach and the only thing what has to be done is to find the asymptotic distribution of the simplicial likelihood depth by using known results on the asymptotic behavior of U-statistics. We demonstrate this for some regression problems. In particular we show that in some cases as regression with exponential distributed errors, which is relevant for reliability theory, the simplicial likelihood depth is not a degenerated U-statistic so that its asymptotic normality follows directly from the Theorem of Hoeffding. Hence there is the hope that in other cases, the simplicial likelihood depth is not a degenerated U-statistic as well.

However, in the most regression problems, the simplicial likelihood depth is a degenerated U-statistic as Liu's simplicial depth. But we demonstrate that this can be treated as well. In these cases the asymptotic distribution is given by the asymptotic distribution of the second term of the Hoeffding decomposition which can be found by the spectral decomposition of the reduced normalized kernel function. For some cases like linear regression through the origin, the spectral decomposition is easy to find.

For other cases like polynomial regression of higher order, this is not so easy but can be done as well. In Section 4 it is shown how this can be done by solving differential equations. For that we derive at first a general formula of the reduced normalized kernel function for general polynomial regression. Then we demonstrate how the spectral decomposition can be found for two cases, namely linear regression with constant term and quadratic regression with constant term. Although the results are derived only for special regression problems we believe that the method can be applied also for other problems.

Section 5 contains a short conclusion and some open problems. The proofs are given in Section 6.

2 Likelihood depth and simplicial likelihood depth

If the variable $Z_n,\ n = 1, \ldots, N$, has a discrete or continuous density function $f_\theta(z_n)$, then let $L(\theta, z_n) = f_\theta(z_n)$ denote the likelihood function at the parameter θ and the observation z_n. We assume that $Z_1, \ldots, Z_N$ are independent and identically distributed throughout the paper. The following definition generalizes the concept of the nonfit of Rousseeuw and Hubert (1999).

Definition 1 (Likelihood nonfit) *$\theta \in I\!R^q$ is a likelihood nonfit within $z_1, \ldots, z_N$ if there is a $\theta' \neq \theta$ with*

$$L(\theta', z_n) > L(\theta, z_n) \text{ for all } n = 1, \ldots, N.$$

A likelihood nonfit θ is also called *not weakly optimal* (Mizera 2002) and was extended as above also in Mizera and Müller (2003). Having the definition of a nonfit the depth of a parameter θ can be defined as in Rousseeuw and Hubert (1999) and Mizera (2002).

Definition 2 (Global likelihood depth) *The global likelihood depth of θ within $z_1, \ldots, z_N$ is the minimal number m of observations $z_{i_1}, \ldots, z_{i_m}$ so that θ is a likelihood nonfit within $\{z_1, \ldots, z_N\} \setminus \{z_{i_1}, \ldots, z_{i_m}\}$.*

Assuming differentiability of the logarithm of the likelihood function $h_n(\theta) = \log L(\theta, z_n)$, a sufficient condition for a likelihood nonfit is

$$u^\top h_n'(\theta) > 0 \text{ for all } n = 1, \ldots, N,$$

for some direction $u \in I\!R^q$, where $h_n'(\theta)$ is the vector of partial first order derivatives at θ. This sufficient condition leads as in Mizera (2002) to tangent likelihood depth, a notion of depth which is more operational (see also Mizera and Müller 2003).

Definition 3 (Tangent likelihood depth) *The tangent likelihood depth $d_T(\theta, z)$ of θ within $z = (z_1, \ldots, z_N)$ is defined as*

$$d_T(\theta, z) := \inf_{u \neq 0} \sharp\{n;\ u^\top h_n'(\theta) \leq 0\}.$$

Since the tangent likelihood depth is more tractable we will work only with this and denote it shortly *likelihood depth*.

Definition 4 (Likelihood depth estimator) $\widehat{\theta}(z)$ *is called a likelihood depth estimator at* z *if*

$$\widehat{\theta}(z) \in \arg\max_{\theta} d_T(\theta, z).$$

Instead of defining the global likelihood depth via the definition of likelihood nonfit, we could define it via a definition of admissibility. A parameter $\theta \in I\!R^q$ would be called admissible if there is no $\theta' \neq \theta$ with

$$L(\theta', z_n) \geq L(\theta, z_n) \text{ for all } n = 1, \ldots, N,$$
$$L(\theta', z_n) > L(\theta, z_n) \text{ for at least one } n = 1, \ldots, N.$$

Then an alternative definition of the global likelihood depth of θ will be the minimal number m of observations $z_{i_1}, \ldots, z_{i_m}$ so that θ is not admissible within $\{z_1, \ldots, z_N\} \setminus \{z_{i_1}, \ldots, z_{i_m}\}$. However, the definition of the tangent likelihood depth would become more complicated, namely like that

$$N - \sup_{u \neq 0}\{\sharp E;\ E \subset \{1, \ldots, N\},\ u^\top h_n'(\theta) \geq 0 \text{ for all } n \in E \text{ and } u^\top h_n'(\theta) > 0 \text{ for at least one } n \in E\}. \tag{1}$$

This is the reason that we prefer the definition via the nonfit although the characterization of the simplicial likelihood depth for polynomial regression in Example 4 would hold in more generality if we would use the definition given by (1).

Example 1 (Multivariate location with elliptical unimodal distribution)
Let $f_\mu : I\!R^q \to I\!R$ be a continuous density satisfying $f_\mu(z_n) = f_0(z_n - \mu)$ so that $\mu \in I\!R^q$ is a q-dimensional location parameter. If there exists a strictly decreasing function $g_0 : [0, \infty) :\to [0, \infty)$ and positive definite matrix Σ with $f_0(v) = g_0(v^\top \Sigma^{-1} v)$ for all $v \in I\!R^q$, then we have $f_0'(v) = g_0'(v^\top \Sigma^{-1} v) 2\, \Sigma^{-1} v$ with $g_0'(v^\top \Sigma^{-1} v) < 0$. This means that f_μ has an elliptical unimodal density. Examples of these densities are the multivariate normal distribution and multivariate Cauchy distribution. For such densities we have

$$h_n'(\mu) = \frac{-g_0'((z_n - \mu)^\top \Sigma^{-1}(z_n - \mu))}{g_0((z_n - \mu)^\top \Sigma^{-1}(z_n - \mu))}\, 2\, \Sigma^{-1}(z_n - \mu)),$$

so that the tangent likelihood depth is

$$d_T(\mu, z) = \inf_{u \neq 0} \sharp\{n :\ u^\top (z_n - \mu) \leq 0\} = \inf_{u \neq 0} \sharp\{n :\ u^\top z_n \leq u^\top \mu\}.$$

But this is the half space depth of Tukey (1975) (see also Donoho and Gasko 1992). Hence Tukey's half space depth is a likelihood depth for any elliptical unimodal distribution.□

Example 2 (Regression with symmetric and unimodal distribution)
Regard a general linear regression model with observations $Z_n = (Y_n, T_n)$ where $Y_n \in I\!R$ is

the dependent variable and $T_n \in I\!R^r$ the independent (explanatory) variable. A common assumption is that the conditional distribution of the dependent variable Y_n given T_n has a symmetric and unimodal density of the form $f_{\beta|T_n=t_n}(y_n) = f_0(y_n - x(t_n)^\top \beta)$, where $\beta \in I\!R^q$ is the unknown parameter and $x : I\!R^r \to I\!R^q$ is a known regression function. As in Example 1 the symmetry and unimodality of $f_{\beta|T_n=t_n}$ means that $f_0(v) = g_0(v^2)$ for all $v \in I\!R$, where g_0 is a strictly decreasing function. Hence we have

$$h'_n(\beta) = \frac{-g'_0((y_n - x(t_n)^\top \beta)^2)}{g_0((y_n - x(t_n)^\top \beta)^2)}\, 2\, \left(y_n - x(t_n)^\top \beta\right) x(t_n),$$

where $g'_0(z) < 0$, so that the tangent likelihood depth is

$$d_T(\beta, z) = \inf_{u \neq 0} \sharp\{n : \left(y_n - x(t_n)^\top \beta\right) u^\top x(t_n) \leq 0\}.$$

This is the regression depth introduced by Rousseeuw and Hubert (1999). Hence this depth concept is a likelihood depth for any symmetric and unimodal distribution. Note that the depth notion proposed in Van Aelst et al. (2002) for polynomial regression, i.e. $x(t_n) = (1, t_n, t_n^2, \ldots, t_n^{q-1})^\top$, is not a likelihood depth. □

In a generalized linear model, we have independent explanatory variables T_n and observations Y_n which depend on T_n. Usually it is assumed that the conditional distribution of Y_n given $T_n = t_n$ is a member of the one-parameter exponential family, i.e. its density is given by

$$f_{\beta|T_n=t_n}(y_n) = \frac{h(y_n)}{c(x(t_n)^\top \beta)} \exp\left(H(y_n)\, q\left(x(t_n)^\top \beta\right)\right),$$

where $h, H, c, q : I\!R \to I\!R$ and $x : I\!R^r \to I\!R^q$ are known functions and $\beta \in I\!R^q$ is the unknown parameter. Because of

$$h'_n(\beta) = \left(\frac{-c'(x(t_n)^\top \beta)}{c(x(t_n)^\top \beta)} + H(y_n)\, q'\left(x(t_n)^\top \beta\right)\right) x(t_n),$$

the tangent likelihood depth is given by

$$d_T(\beta, z) = \inf_{u \neq 0} \sharp\{n : \left(H(y_n)\, q'\left(x(t_n)^\top \beta\right) - \frac{c'(x(t_n)^\top \beta)}{c(x(t_n)^\top \beta)}\right) u^\top x(t_n) \leq 0\}. \tag{2}$$

Under special assumptions on q and c we have the following characterization of this likelihood depth for generalized linear models.

Theorem 1 *If $q'(v) > 0$ for all $v \in I\!R$ and b given by $b(v) = \frac{c'(v)}{c(v)\, q'(v)}$ is strictly decreasing or strictly increasing so that b^{-1} exists, then the tangent likelihood depth for a generalized linear model is given by*

$$d_T(\beta, z) = \inf_{u \neq 0} \sharp\{n : \left(b^{-1}(H(y_n)) - x(t_n)^\top \beta\right) u^\top x(t_n) \leq 0\}, \tag{3}$$

i.e. the tangent likelihood depth is the regression depth of Rousseeuw and Hubert (1999) for the transformed observations $b^{-1}(H(y_n))$.

The equivalence of (2) and (3) is obvious from the assumptions of Theorem 1. Note that this is based on the monotone invariance property of regression depth shown by Proposition 2 of Van Aelst et al. (2002). But, if the assumptions of Theorem 1 are not satisfied, then the likelihood depth for a generalized linear model can lead to a new depth notion which cannot be interpreted as the regression depth of Rousseeuw and Hubert (1999). However, the most well known generalized linear models satisfy the assumptions of Theorem 1:

Example 3 (Examples of generalized linear models)
For regression with exponential distributed dependent observations Y_n, the density of the conditional distribution of Y_n has the form $f_{\beta|T_n=t_n}(y_n) = \lambda_n \exp(-\lambda_n y_n)$ with $\lambda_n = \exp(-x(t_n)^\top \beta)$. Then we have $H(y_n) = y_n$, $q(v) = -\exp(-v)$, $h(y_n) = 1$, $c(v) = \exp(v)$ such that $b(v) = \exp(v)$ and $b^{-1}(H(y_n)) = \log(y_n)$. We get the same $b(v)$ and $b^{-1}(H(y_n))$ for a loglinear model, where the dependent observations Y_n have a Poisson distribution with $f_{\beta|T_n=t_n}(y_n) = \frac{\lambda_n^{y_n} \exp(-\lambda_n)}{y_n!}$ and $\lambda_n = \exp(x(t_n)^\top \beta)$, so that $H(y_n) = y_n$, $q(v) = v$, $h(y_n) = \frac{1}{y_n!}$, $c(v) = \exp(\exp(v))$. $b^{-1}(H(y_n)) = \log(y_n)$ holds also for observations with geometrical distribution since $f_{\beta|T_n=t_n}(y_n) = p_n(1-p_n)^{y_n}$ with $p_n = \frac{1}{1+\exp(x(t_n)^\top \beta)}$ implies $H(y_n) = y_n$, $q(v) = \log\left(\frac{\exp(v)}{1+\exp(v)}\right)$, $h(y_n) = 1$, $c(v) = 1 + \exp(v)$. However, for logistic regression, where the dependent observations have a binomial distribution, the observations have to be transformed differently. In this case we have $f_{\beta|T_n=t_n}(y_n) = \binom{m_n}{y_n} p_n^{y_n}(1-p_n)^{m_n-y_n}$ with $p_n = F(x(t_n)^\top \beta)$, where $F(v) = \frac{\exp(v)}{1+\exp(v)}$ is the logistic function. Here we have $H(y_n) = y_n$, $q(v) = v$, $h(y_n) = \binom{m_n}{y_n}$, $c(v) = (1+\exp(v))^{m_n}$, so that $b(v) = m_n\, F(v)$ and $b^{-1}(H(y_n)) = F^{-1}\left(\frac{y_n}{m_n}\right)$. In the special case of Bernoulli distribution, i.e. $y_n \in \{0,1\}$ and $m_n = 1$, the resulting likelihood depth coincides with the overlap measure of Christmann and Rousseeuw (2001). □

In all examples for generalized linear models, it turned out that the likelihood depth coincides with the regression depth of Rousseeuw and Hubert (1999) if the dependent observations are appropriately transformed. This means that the likelihood depth for these generalized linear models has the same robustness properties as shown by Rousseeuw and Hubert (1999) and Van Aelst and Rousseeuw (2000) for regression depth. In particular the likelihood depth estimator has a breakdown point of $\frac{1}{3}$ for multiple regression, i.e. for $x(t_n) = (1, t_n)^\top$ with $t_n \in I\!R^r$. Moreover, likelihood depth and the likelihood depth estimator can be calculated by the methods proposed by Rousseeuw and Hubert (1999), Rousseeuw and Struyf (1998), Van Aelst et al. (2002).

The calculation of the likelihood depth is in particular easy for $q + 1$ observations. Counting all subsets with $q + 1$ observations which has a likelihood depth greater than zero leads to the simplicial likelihood depth.

Definition 5 (Simplicial likelihood depth) *If d_T is a tangent likelihood depth for $\theta \in I\!R^q$, then the simplicial likelihood depth $d_S(\theta, z)$ of θ within $z = (z_1, \ldots, z_N)$ is defined as*

$$\begin{aligned} d_S(\theta, z) &:= \binom{N}{q+1}^{-1} \sharp \left\{ \{n_1, \ldots, n_{q+1}\} \subset \{1, \ldots, N\};\ d_T\left(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})\right) > 0 \right\} \\ &= \binom{N}{q+1}^{-1} \sum_{\substack{n_1, \ldots, n_{q+1} \\ \text{pairwise different}}} 1\left\{ d_T(\theta, (z_{n_1}, \ldots, z_{n_{q+1}})) > 0 \right\}, \end{aligned}$$

where $1\{d_T(\theta, (z_1, \ldots, z_{q+1})) > 0\}$ denotes the indicator function

$$1_{\{d_T(\theta,(z_1,\ldots,z_{q+1}))>0\}}((z_1, \ldots, z_{q+1})).$$

The name of this depth criterion is motivated by the example for multivariate location. In Example 1 it was shown that the tangent likelihood depth for multivariate location $\mu \in I\!R^q$ with elliptical and unimodal distribution is Tukey's half space depth. This half space depth satisfies $d_T\left(\mu, (z_{n_1}, \ldots, z_{n_{q+1}})\right) > 0$ if and only if μ lies in the simplex spanned by $z_{n_1}, \ldots, z_{n_{q+1}}$. Hence the simplicial likelihood depth is counting the simplices which contain μ. But this is the simplicial depth introduced by Liu (1988, 1990).

Example 4 (Regression)
The Examples 2 and 3 have shown that the tangent likelihood depths for the most common regression models coincide with the regression depth of Rousseeuw and Hubert (1999) after an appropriate transformation of the dependent observations. W.l.o.g. let be $y_1, \ldots, y_N$ the appropriately transformed observations. Then the simplicial likelihood depth is counting all subsets $z_{n_1}, \ldots, z_{n_{q+1}}$ with $\inf_{u \neq 0} \sharp\{i :\ (y_{n_i} - x(t_{n_i})^\top \beta)\ u^\top x(t_{n_i}) \leq 0\} > 0$. For polynomial regression with $x(t_n) = (1, t_n, t_n^2, \ldots, t_n^{q-1})^\top$ and $t_{n_1} < t_{n_2} < \ldots < t_{n_{q+1}}$, we have $\inf_{u \neq 0} \sharp\{i :\ (y_{n_i} - x(t_{n_i})^\top \beta)\ u^\top x(t_{n_i}) \leq 0\} > 0$ if

$$(y_{n_i} - x(t_{n_i})^\top \beta)\,(-1)^i \geq 0 \text{ for all } i = n_1, \ldots, n_{q+1}$$

or

$$(y_{n_i} - x(t_{n_i})^\top \beta)\,(-1)^i \leq 0 \text{ for all } i = n_1, \ldots, n_{q+1},$$

i.e. the residuals have alternating signs. This condition is also necessary with probability one if Y_n has a continuous distribution since in this case $Y_n - x(t_n)^\top \beta \neq 0$ with probability one. If we generally assume that $t_1 < t_2 < \ldots < t_N$, then the simplicial likelihood depth for polynomial regression with continuous Y_n is given with probability one by

$$\begin{aligned} d_S(\beta, z) &= \binom{N}{q+1}^{-1} \sum_{n_1 < n_2 < \ldots < n_{q+1}} \left(\prod_{i=1}^{q+1} 1\left\{ (y_{n_i} - x(t_{n_i})^\top \beta)\ (-1)^i \geq 0 \right\} \right. \qquad (4) \\ &\quad \left. + \prod_{i=1}^{q+1} 1\left\{ (y_{n_i} - x(t_{n_i})^\top \beta)\ (-1)^i \leq 0 \right\} \right), \end{aligned}$$

where $1\left\{(y_n - x(t_n)^\top \beta)\,(-1)^i \geq 0\right\}$ is an abbreviation for the indicator function $1_{\left\{(y_n - x(t_n)^\top \beta)\,(-1)^i \geq 0\right\}}((y_n, t_n))$. For linear regression this notion of simplicial depth for regression were derived in Rousseeuw and Hubert (1999) via the dual approach. Here we get the property (4) only with probability one for continuous Y_n. But if we would define the tangent likelihood depth via admissibility, i.e. by (1), the characterization (4) would hold always and for any distribution of Y_n. Note that if we would base the simplicial depth on the depth notion proposed by Van Aelst et al. (2002) for polynomial regression, we would not get the depth function (4). □

3 Tests based on the simplicial likelihood depth

For very small sample sizes, the distribution of the simplicial likelihood depth $d_S(\theta, Z)$ under θ can be calculated by combinatorial methods. However, for large data sets, approximations of the distribution are necessary. For that note that the tangent likelihood depth is a symmetric kernel, i.e. it satisfies $d_T(\theta, (z_1, \ldots, z_N)) = d_T(\theta, (z_{\pi(1)}, \ldots, z_{\pi(N)}))$ for all permutations $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$. Hence the simplicial likelihood depth is a U-statistic with symmetric kernel function

$$\psi_\theta(z_1, \ldots, z_{q+1}) = 1\left\{d_T\left(\theta, (z_1, \ldots, z_{q+1})\right) > 0)\right\}.$$

The asymptotic distribution of U-statistics is well known. In particular if the U-statistic is not degenerated, i.e. $\psi_\theta^1(z_1) := E(\psi_\theta(Z_1, \ldots, Z_{q+1})|Z_1 = z_1)$ is not independent of z_1, then we have with the Theorem of Hoeffding (see e.g. Lee 1990, p. 76, or Witting and Müller-Funk 1995, p. 635)

$$\mathcal{L}(\sqrt{N}(d_S(\theta, (Z_1, \ldots, Z_N)) - \gamma_\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (q+1)^2\,\sigma_\theta^2)$$

with $\gamma_\theta = E(\psi_\theta(Z_1, \ldots, Z_{q+1}))$ and $\sigma_\theta^2 = \mathrm{Var}(\psi_\theta^1(Z_1))$. Hence a test for testing $H_0 : \theta \in \Theta_0$ against $H_0 : \theta \notin \Theta_0$, where Θ_0 is a subset of the parameter space, can be based on the test statistic $T(z_1, \ldots, z_N) := \sup_{\theta \in \Theta_0} T_\theta(z_1, \ldots, z_N)$ where

$$T_\theta(z_1, \ldots, z_N) := \frac{\sqrt{N}(d_S(\theta, (z_1, \ldots, z_N)) - \gamma_\theta)}{(q+1)\;\sigma_\theta}. \tag{5}$$

If the null hypothesis H_0 is rejected if $T(z_1, \ldots, z_N)$ is less than the α-quantile of the standard normal distribution then this test is asymptotically an α-level test since for any $c \in I\!R$ and all $\theta \in \Theta_0$

$$P_{\theta_0}\left(\sup_{\theta \in \Theta_0} T_\theta(z_1, \ldots, z_N) < c\right) \leq P_{\theta_0}\left(T_{\theta_0}(z_1, \ldots, z_N) < c\right).$$

We will see later that usually the quantities γ_θ and σ_θ are independent of θ so that the test has a very simple form. The main difficulty is the calculation of $\sup_{\theta \in \Theta_0} d_S(\theta, (z_1, \ldots, z_N))$.

This difficulty disappear for tests of $H_0 : \theta = \theta_0$ against $H_0 : \theta \neq \theta_0$ where θ_0 is a given parameter. These tests also can be used to create confidence regions by defining the confidence regions as the set of all parameters θ_0 for which $H_0 : \theta = \theta_0$ is not rejected.

Unfortunately, the simplicial likelihood depth is a degenerated U-statistic in many cases. This is not only the case for Liu's (1988, 1990) simplicial depth for multivariate location. For polynomial regression (see Example 4) it depends whether $P(Y_n - x(T_n)^\top \beta \geq 0|T_n)$ is equal $\frac{1}{2}$ or not. To see this let

$$\mathcal{P}_{q+1} := \{\pi : \{1, \ldots, q+1\} \to \{1, \ldots, q+1\};\ \pi(i) \neq \pi(j) \text{ for } i \neq j\}$$

the set of all permutations of $\{1, \ldots, q+1\}$. Then the simplicial likelihood depth for polynomial regression can be written as (compare with Example 4)

$$d_S(\beta, z) = \binom{N}{q+1}^{-1} \sum_{\substack{n_1, \ldots, n_{q+1} \\ \text{pairwise different}}} \psi_\beta(z_{n_1}, \ldots, z_{n_{q+1}})$$

where

$$\psi_\beta(z_1, \ldots, z_{q+1}) = \sum_{\pi \in \mathcal{P}_{q+1}} \left(\prod_{i=1}^{q} 1\left\{ t_{\pi(i)} < t_{\pi(i+1)} \right\} \right) \tag{6}$$
$$\cdot \left(\prod_{i=1}^{q+1} 1\left\{ (y_{\pi(i)} - x(t_{\pi(i)})^\top \beta)\ (-1)^i \geq 0 \right\} + \prod_{i=1}^{q+1} 1\left\{ (y_{\pi(i)} - x(t_{\pi(i)})^\top \beta)\ (-1)^i \leq 0 \right\} \right).$$

Then ψ_β is a symmetric kernel function.

Proposition 1 *Let be $p = P(Y_n - x(T_n)^\top \beta \geq 0|T_n)$ with probability 1 and T_n has an absolute continuous distribution.*
a) If $q+1$ is even, then with probability 1

$$\begin{aligned} & E\left(\psi_\beta((Z_1, \ldots, Z_{q+1})|Z_1 = (y_1, t_1))\right) \\ = \ & p^{\frac{q+1}{2}-1} (1-p)^{\frac{q+1}{2}-1} \cdot \left((1-p)\, 1\{y_1 - x(t_1)^\top \beta \geq 0\} + p\, 1\{y_1 - x(t_1)^\top \beta \leq 0\} \right). \end{aligned}$$

b) If $q+1$ is odd, then with probability 1

$$\begin{aligned} & E\left(\psi_\beta((Z_1, \ldots, Z_{q+1})|Z_1 = (y_1, t_1))\right) \\ = \ & p^{\frac{q}{2}-1} (1-p)^{\frac{q}{2}-1}\, q! \cdot \Bigg(p\,(1-p) \sum_{m=0}^{\frac{q}{2}} P(T_1 < T_2 < \ldots < T_{q+1}|T_{2m+1} = t_1) \\ & + \left((1-p)^2\, 1\{y_1 - x(t_1)^\top \beta \geq 0\} + p^2\, 1\{y_1 - x(t_1)^\top \beta \leq 0\} \right) \\ & \cdot \sum_{m=1}^{\frac{q}{2}} P(T_1 < T_2 < \ldots < T_{q+1}|T_{2m} = t_1) \Bigg). \end{aligned}$$

c) If $p = \frac{1}{2}$, then $E(\psi_\beta((Z_1, \ldots, Z_{q+1})|Z_1 = (y_1, t_1))) = \left(\frac{1}{2}\right)^q$ with probability 1.

Proposition 1 shows that the simplicial likelihood depth for polynomial regression is a degenerated U-statistic if the conditional probability of nonnegative residuals given T_n is exactly $\frac{1}{2}$. This is the case if the median of the conditional residual distribution is zero, a case which is often satisfied. However, models exist as well where the median of the residual distribution is not zero. This is for example the case for exponential distributed dependent observations as the following example shows.

Example 5 (Regression with exponential distribution)
As in the first example of Example 3 we regard a regression experiment where the dependent variables Y_n possesses an exponential distribution. Replacing the dependent observations by the logarithm of their values and using the parametrization $\lambda_n = \exp(-x(t_n)^\top \beta)$ the simplicial likelihood depth for polynomial regression is given by (4). In particular it is a U-statistic with kernel function

$$\begin{aligned}\psi_\beta(z_1, \ldots, z_{q+1}) = \sum_{\pi \in \mathcal{P}_{q+1}} \left(\prod_{i=1}^{q} 1\left\{ t_{\pi(i)} < t_{\pi(i+1)} \right\} \right) \\ \cdot \left(\prod_{i=1}^{q+1} 1\left\{ (\log(y_{\pi(i)}) - x(t_{\pi(i)})^\top \beta)\ (-1)^i \geq 0 \right\} \right. \\ \left. + \prod_{i=1}^{q+1} 1\left\{ (\log(y_{\pi(i)}) - x(t_{\pi(i)})^\top \beta)\ (-1)^i \leq 0 \right\} \right).\end{aligned}$$

Hence the quantity p of Proposition 1 is

$$\begin{aligned} p &= P(\log(Y_n) - x(T_n)^\top \beta \geq 0 | T_n = t_n) = P(Y_n - \exp(x(T_n)^\top \beta) \geq 0 | T_n = t_n) \\ &= P_{\lambda_n}\left(Y_n \geq \frac{1}{\lambda_n} \right) = \frac{1}{e} \neq \frac{1}{2}, \end{aligned}$$

so that the simplicial likelihood depth is a nondegenerated U-statistic. For deriving the test statistic (5), we have only to calculate γ_β and σ_β.

The calculation of γ_β and σ_β is demonstrated here for simple linear regression, i.e. for $x(t_n) = (1, t_n)^\top$ and $q + 1 = 3$. Again we assume that G is a differentiable distribution function of the explanatory variables. At first note that Lemma 3 provides

$$\begin{aligned} & P(T_1 < T_2 < T_3 | T_1 = t_1) + P(T_1 < T_2 < T_3 | T_3 = t_1) \\ &= \frac{1}{2} - G(t_1) + \frac{1}{2}\, G(t_1)^2 + \frac{1}{2}\, G(t_1)^2 = \frac{1}{2} - G(t_1) + G(t_1)^2, \end{aligned}$$

and

$$P(T_1 < T_2 < T_3 | T_2 = t_1) = G(t_1) - G(t_1)^2.$$

Then we have

$$
\begin{aligned}
&\psi_\beta^1((y_1,t_1)) \\
&= 2\left(1-\frac{1}{e}\right)\frac{1}{e}\left(\frac{1}{2}-G(t_1)+G(t_1)^2\right) \\
&\quad + 2\left(\left(1-\frac{1}{e}\right)^2 1\{\log(y_1)-x(t_1)^\top\beta\geq 0\}+\left(\frac{1}{e}\right)^2 1\{\log(y_1)-x(t_1)^\top\beta\leq 0\}\right) \\
&\qquad \cdot\left(G(t_1)-G(t_1)^2\right).
\end{aligned}
$$

Using $\int_a^b G(x)^k\, g(x)\, dx = \frac{1}{k+1}\left(G(b)^{k+1}-G(a)^{k+1}\right)$ for $g = G'$, we obtain

$$
\begin{aligned}
\gamma_\beta &= E(\psi_\beta^1((Y_1,T_1))) \\
&= \left(1-\frac{1}{e}\right)\frac{1}{e}\frac{2}{3}+2\left(1-\frac{1}{e}\right)\frac{1}{e}\left(\frac{1}{2}-\frac{1}{3}\right)=\left(1-\frac{1}{e}\right)\frac{1}{e}=0.2325,
\end{aligned}
$$

$$
\begin{aligned}
&E\left(\psi_\beta^1((Y_1,T_1))^2\right) \\
&= E\left(4\left(1-\frac{1}{e}\right)^2\left(\frac{1}{e}\right)^2\left(\frac{1}{4}-G(T_1)+2G(T_1)^2-2G(T_1)^3+G(T_1)^4\right)\right) \\
&\quad + E\left(4\left(\left(1-\frac{1}{e}\right)^4 1\{\log(y_1)-x(t_1)^\top\beta\geq 0\}+\left(\frac{1}{e}\right)^4 1\{\log(y_1)-x(t_1)^\top\beta\leq 0\}\right)\right. \\
&\quad \left.\cdot\left(G(T_1)^2-2G(T_1)^3+G(T_1)^4\right)\right) \\
&\quad + E\left(8\left(1-\frac{1}{e}\right)\frac{1}{e}\right. \\
&\quad \cdot\left(\left(1-\frac{1}{e}\right)^2 1\{\log(y_1)-x(t_1)^\top\beta\geq 0\}+\left(\frac{1}{e}\right)^2 1\{\log(y_1)-x(t_1)^\top\beta\leq 0\}\right) \\
&\quad \left.\cdot\left(\frac{1}{2}G(T_1)-\frac{3}{2}G(T_1)^2+2G(T_1)^3-G(T_1)^4\right)\right) \\
&= 4\left(1-\frac{1}{e}\right)^2\left(\frac{1}{e}\right)^2\left(\frac{1}{4}-\frac{1}{2}+\frac{2}{3}-\frac{1}{2}+\frac{1}{5}\right) \\
&\quad + 4\left(1-\frac{1}{e}\right)\frac{1}{e}\left(\left(1-\frac{1}{e}\right)^3+\left(\frac{1}{e}\right)^3\right)\left(\frac{1}{3}-\frac{1}{2}+\frac{1}{5}\right) \\
&\quad + 8\left(1-\frac{1}{e}\right)^2\left(\frac{1}{e}\right)^2\left(\frac{1}{4}-\frac{1}{2}+\frac{1}{2}-\frac{1}{5}\right) \\
&= \left(1-\frac{1}{e}\right)^2\left(\frac{1}{e}\right)^2\left(\frac{7}{15}+\frac{6}{15}\right)+\left(1-\frac{1}{e}\right)\frac{1}{e}\left(\left(1-\frac{1}{e}\right)^3+\left(\frac{1}{e}\right)^3\right)\frac{2}{15},
\end{aligned}
$$

$$\sigma_\beta^2 = \text{Var}\left(\psi_\beta^1((Y_1,T_1))\right)$$
$$= \left(1-\frac{1}{e}\right)^2 \left(\frac{1}{e}\right)^2 \frac{(-2)}{15} + \left(1-\frac{1}{e}\right) \frac{1}{e} \left(\left(1-\frac{1}{e}\right)^3 + \left(\frac{1}{e}\right)^3\right) \frac{2}{15} = 0.0022.$$

Hence the test statistic

$$\frac{\sqrt{N}(d_S(\beta_0,(Z_1,\ldots,Z_N)) - 0.2325)}{0.1396}$$

has approximately a standard normal distribution and can be used for testing hypotheses over β for exponential distributed dependent observations, which in particular appear in reliability theory. Note that these tests do not depend on the distribution of the explanatory variables if T_n has a continuous distribution. Moreover these tests can be used not only for exponential distributed Y_n but also for all continuous distributions which satisfy $P(\log(Y_n) - x(T_n)^\top\beta \geq 0|T_n = t_n) = \frac{1}{e}$ for almost all t_n. □

As soon as $P(Y_n - x(T_n)^\top\beta \geq 0|T_n) = \frac{1}{2}$ holds, and this is the case in the most regression setups, then the simplicial likelihood depth is degenerated. But asymptotic distributions can be also derived for degenerated U-statistics by using the second component of the Hoeffding decomposition. We have namely the following result (see e.g. Lee 1990, p. 79, 80, 90, Witting and Müller-Funk, p. 650). If the reduced normalized kernel function

$$\psi_\theta^2(z_1,z_2) := E(\psi_\theta(Z_1,\ldots,Z_{q+1}) - \gamma_\theta|Z_1 = z_1, Z_2 = z_2)$$

is $I\!\!L_2$-integrable, then it has a spectral decomposition of the form

$$\psi_\theta^2(z_1,z_2) = \sum_{l=1}^{\infty} \lambda_l \varphi_l(z_1)\varphi_l(z_2),$$

where the functions φ_l are $I\!\!L_2$-integrable, normalized, and orthogonal. Then the asymptotic distribution of the simplicial likelihood depth is given by

$$\mathcal{L}(N(d_S(\theta,(Z_1,\ldots,Z_N)) - \gamma_\theta)) \xrightarrow{\mathcal{L}} \mathcal{Q}\left(\binom{q+1}{2}\lambda_l;\ l \in I\!\!N\right), \tag{7}$$

where $\mathcal{Q}(\lambda_l^0;\ l \in I\!\!N)$ is the distribution of the random variable $\sum_{l=1}^\infty \lambda_l^0(X_l^2 - 1)$ with $X_l \sim \mathcal{N}(0,1)$. In the general case, it could happen that the eigenvalues λ_l depend on the underlying parameter θ. But in the examples studied below this is not the case. Also γ_θ is independent of θ there. Having a asymptotic distribution which is independent of θ, tests for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \notin \Theta_0$ can be constructed as in the nondegenerate case as explained above. In particular the test statistic can be based on

$$\sup_{\theta\in\Theta_0} N(d_S(\theta,(z_1,\ldots,z_N)) - \gamma_\theta). \tag{8}$$

In some cases it is simple to find the spectral decomposition of $\psi_\theta^2(z_1,z_2)$. This is for example the case for simple linear regression through the origin as the example below shows. In other cases as for general polynomial regression, the derivation of the spectral decomposition needs more steps. This is demonstrated in the next section.

Example 6 (Linear regression through the origin)
For linear regression through the origin, the regression function satisfies $x(t_n) = t_n \in I\!R$ and the unknown parameter is $\beta \in I\!R$, so that the tangent likelihood depth is $d_T(\beta, z) = \inf_{u \neq 0} \sharp\{n : \ (y_n - t_n\beta)\, ut_n \leq 0\}$. The tangent likelihood depth of two observations (y_1, t_1), (y_2, t_2) is greater than 0 if and only if $(y_1 - t_1\beta)\, t_1$ and $(y_2 - t_2\beta)\, t_2$ have different signs or at least one of them is equal to zero. Hence the simplicial likelihood depth is

$$d_S(\beta, z) = \binom{N}{2}^{-1} \sum_{n_1 \neq n_2} \psi_\beta(z_{n_1}, z_{n_2})$$

with kernel function

$$\begin{aligned} &\psi_\beta(z_1, z_2) \\ &= 1\{(y_1 - t_1\beta)\, t_1 \geq 0\}\, 1\{(y_2 - t_2\beta)\, t_2 \leq 0\} \\ &\quad + 1\{(y_1 - t_1\beta)\, t_1 \leq 0\}\, 1\{(y_2 - t_2\beta)\, t_2 \geq 0\}. \end{aligned}$$

As for regression with intercept treated in Proposition 1, this is a degenerated U-statistic for continuous distributions since

$$\begin{aligned} &\psi_\beta^1(z_1) = E(\psi_\beta(Z_1, Z_2) | Z_1 = z_1) \\ &= 1\{(y_1 - t_1\beta)\, t_1 \geq 0\}\, P((Y_2 - T_2\beta)\, T_2 \leq 0) \\ &\quad + 1\{(y_1 - t_1\beta)\, t_1 \leq 0\}\, P((Y_2 - T_2\beta)\, T_2 \geq 0) \\ &= 1\{(y_1 - t_1\beta)\, t_1 \geq 0\}\, \left(P(Y_2 - T_2\,\beta \leq 0)\, P(T_2 \geq 0) + P(Y_2 - T_2\,\beta \geq 0)\, P(T_2 \leq 0)\right) \\ &\quad + 1\{(y_1 - t_1\beta)\, t_1 \leq 0\}\, \left(P(Y_2 - T_2\,\beta \geq 0)\, P(T_2 \geq 0) + P(Y_2 - T_2\,\beta \leq 0)\, P(T_2 \leq 0)\right) \\ &= \frac{1}{2}. \end{aligned}$$

The spectral decomposition of $\psi_\beta^2(z_1, z_2) = \psi_\beta(z_1, z_2) - \frac{1}{2}$ is

$$\psi_\beta(z_1, z_2) - \frac{1}{2} = -\frac{1}{2}\, \varphi(z_1)\, \varphi(z_2)$$

with

$$\varphi(z_1) = (1\{y_1 - t_1\beta \geq 0\} - 1\{y_1 - t_1\beta \leq 0\})\ (1\{t_1 \geq 0\} - 1\{t_1 \leq 0\})\,.$$

To see this, set $r_i^+ = 1\{y_1 - t_1\beta \geq 0\}$, $r_i^- = 1\{y_1 - t_1\beta \leq 0\}$, $t_i^+ = 1\{t_1 \geq 0\}$, $t_i^- = 1\{t_1 \leq 0\}$ for $i = 1, 2$ and $p_r = (r_1^+ - r_1^-)\,(r_2^+ - r_2^-)$, $p_t = (t_1^+ - t_1^-)\,(t_2^+ - t_2^-)$. Then we have

$$\begin{aligned} p_r &= 2\, r_1^+\, r_2^+ + 2\, r_1^-\, r_2^- - 1 = 1 - 2\, r_1^+\, r_2^- - 2\, r_1^-\, r_2^+, \\ p_t &= 2\, t_1^+\, t_2^+ + 2\, t_1^-\, t_2^- - 1 = 1 - 2\, t_1^+\, t_2^- - 2\, t_1^-\, t_2^+, \end{aligned}$$

so that

$$\begin{aligned}
&\psi_\beta(z_1,z_2)\\
&= r_1^+\,t_1^+\,r_2^-\,t_2^+ + r_1^+\,t_1^+\,r_2^+\,t_2^- + r_1^-\,t_1^-\,r_2^-\,t_2^+ + r_1^-\,t_1^-\,r_2^+\,t_2^-\\
&\quad + r_1^-\,t_1^+\,r_2^+\,t_2^+ + r_1^-\,t_1^+\,r_2^-\,t_2^- + r_1^+\,t_1^-\,r_2^+\,t_2^+ + r_1^+\,t_1^-\,r_2^-\,t_2^-\\
&= t_1^+\,t_2^+\,\left(r_1^+\,r_2^- + r_1^-\,r_2^+\right) + t_1^-\,t_2^-\,\left(r_1^+\,r_2^- + r_1^-\,r_2^+\right)\\
&\quad + t_1^+\,t_2^-\,\left(r_1^+\,r_2^+ + r_1^-\,r_2^-\right) + t_1^-\,t_2^+\,\left(r_1^+\,r_2^+ + r_1^-\,r_2^-\right)\\
&= t_1^+\,t_2^+\,\left(\frac{1}{2}-\frac{p_r}{2}\right) + t_1^-\,t_2^-\,\left(\frac{1}{2}-\frac{p_r}{2}\right)\\
&\quad + t_1^+\,t_2^-\,\left(\frac{1}{2}+\frac{p_r}{2}\right) + t_1^-\,t_2^+\,\left(\frac{1}{2}+\frac{p_r}{2}\right)\\
&= \left(\frac{1}{2}-\frac{p_r}{2}\right)\left(\frac{1}{2}+\frac{p_t}{2}\right) + \left(\frac{1}{2}+\frac{p_r}{2}\right)\left(\frac{1}{2}-\frac{p_t}{2}\right)\\
&= \frac{1}{2}-\frac{p_r\,p_t}{2}.
\end{aligned}$$

Hence the test statistic

$$N\left(d_S(\beta_0,(Z_1,\ldots,Z_N))-\frac{1}{2}\right)$$

has approximately a $\mathcal{Q}\left(-\frac{1}{2}\right)$ distribution. An α-quantile $q_{\mathcal{Q}\left(-\frac{1}{2}\right)}(\alpha)$ of this distribution satisfies $q_{\mathcal{Q}\left(-\frac{1}{2}\right)}(\alpha) = \frac{1}{2}\left(1-q_{\chi^2}(1-\alpha)\right)$, where $q_{\chi^2}(\alpha)$ is the α-quantile of the central χ^2-distribution. Note that this test is a distribution free tests if (Y_n,T_n) has a continuous distribution with $P(Y_n-T_N\geq 0|T_n)=\frac{1}{2}$.□

4 Polynomial regression

Throughout this section, we assume a polynomial regression model with $P(Y_n - x(T_n)^\top\beta \geq 0|T_n) = \frac{1}{2}$ and differentiable distribution function G of the distribution of T_n. In particular we have $x(t_n) = (1, t_n, t_n^2, \ldots, t_n^{q-1})^\top$ and $\beta \in I\!R^q$. The kernel function $\psi_\beta(z_1,\ldots,z_{q+1})$ of the simplicial likelihood depth is given by (6). We know from Proposition 1 c) that $\gamma_\beta = E(\psi_\beta(Z_1,\ldots,Z_{q+1})) = E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1) = \left(\frac{1}{2}\right)^q$ so that the simplicial likelihood depth is a degenerated U-statistic. The first step for deriving the asymptotic distribution of the simplicial likelihood depth is to calculate the reduced normalized kernel-function $\psi_\beta^2(z_1,z_2) = E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,Z_2=z_2) - \left(\frac{1}{2}\right)^q$. Set $r_n := y_n - x(t_n)^\top\beta$ and

$$\tau(r_1,r_2) := 1\{r_1\geq 0\}\;1\{r_2\leq 0\} + 1\{r_1\leq 0\}\;1\{r_2\geq 0\}.$$

Proposition 2 *With probability 1, we have*

$$\begin{aligned} & E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,Z_2=z_2) - \left(\frac{1}{2}\right)^q \\ = & \left(\tau(r_1,r_2)-\frac{1}{2}\right)\left(\frac{1}{2}-|G(t_1)-G(t_2)|\right)^{q-1}. \end{aligned}$$

We obtain in particular for linear regression ($q = 2$)

$$\begin{aligned} & E(\psi_\beta(Z_1,Z_2,Z_3)|Z_1=z_1,Z_2=z_2) - \frac{1}{4} \\ = & \left(\tau(r_1,r_2)-\frac{1}{2}\right)\left(\frac{1}{2}-|G(t_1)-G(t_2)|\right), \end{aligned} \tag{9}$$

and for quadratic regression ($q = 3$)

$$\begin{aligned} & E(\psi_\beta(Z_1,Z_2,Z_3,Z_4)|Z_1=z_1,Z_2=z_2) - \frac{1}{8} \\ = & \left(\tau(r_1,r_2)-\frac{1}{2}\right)\left(\frac{1}{4}-|G(t_1)-G(t_2)|+(G(t_1)-G(t_2))^2\right). \end{aligned} \tag{10}$$

For these two cases we will now demonstrate how the singular value decomposition can be found. At first it is easy to see (compare also with Example 6) that the spectral decomposition of $\left(\tau(r_1,r_2)-\frac{1}{2}\right)$ is

$$\tau(r_1,r_2)-\frac{1}{2} = -\frac{1}{2}\varphi_*(r_1)\varphi_*(r_2) \tag{11}$$

with $\varphi_*(r) = 1\{r \le 0\} - 1\{r \ge 0\}$. Hence we need only to find the spectral decomposition of $\frac{1}{2}-|G(t_1)-G(t_2)|$ and $\frac{1}{4}-|G(t_1)-G(t_2)|+(G(t_1)-G(t_2))^2$. But this can be done by finding the spectral decomposition of $\frac{1}{2}-|t_1-t_2|$ and $\frac{1}{4}-|t_1-t_2|+(t_1-t_2)^2$ for the uniform distribution on $[0,1]$ since substitution provides

$$\begin{aligned} 0 &= \int_0^1 \varphi_l(t)\,\varphi_k(t)\,dt = \int_{-\infty}^{\infty} \varphi_l(G(t))\,\varphi_k(G(t))\,g(t)\,dt, \\ 1 &= \int_0^1 \varphi_l(t)^2\,dt = \int_{-\infty}^{\infty} \varphi_l(G(t))^2\,g(t)\,dt, \end{aligned}$$

where $g(t) = G'(t)$. To find the spectral decomposition of $\frac{1}{2}-|t_1-t_2|$ and $\frac{1}{4}-|t_1-t_2|+(t_1-t_2)^2$ we calculate the eigenvalues and the eigenfunctions by setting

$$\lambda\,\varphi(s) = \int_0^1 \left(\frac{1}{2}-|t-s|\right)\,\varphi(t)\,dt$$

and

$$\lambda\,\varphi(s) = \int_0^1 \left(\frac{1}{4}-|t-s|+(t-s)^2\right)\,\varphi(t)\,dt,$$

respectively. Differentiation of these equations leads to differential equations whose solutions provides candidates of the eigenfunctions.

Lemma 1 *The spectral decomposition of* $\frac{1}{2} - |t-s|$ *is given by*

$$\frac{1}{2} - |t-s| = \sum_{l=1}^{\infty} \lambda_l \; \varphi_l(t) \; \varphi_l(s),$$

where

$$\lambda_{2l-1} = \frac{2}{\pi^2 \; (2l-1)^2}, \quad \varphi_{2l-1}(t) = \sqrt{2} \; \cos((2l-1)\,\pi\, t),$$
$$\lambda_{2l} = \frac{2}{\pi^2 \; (2l-1)^2}, \quad \varphi_{2l}(t) = \sqrt{2} \; \sin((2l-1)\,\pi\, t)$$

for $l \in I\!N$.

Lemma 2 *The spectral decomposition of* $\frac{1}{4} - |t-s| + (t-s)^2$ *is given by*

$$\frac{1}{4} - |t-s| + (t-s)^2 = \sum_{l=0}^{\infty} \lambda_l \; \varphi_l(t) \; \varphi_l(s),$$

where

$$\lambda_0 = \frac{1}{12}, \quad \varphi_0 = 1,$$
$$\lambda_{2l-1} = \frac{2}{\pi^2 \; (2l)^2}, \quad \varphi_{2l-1}(t) = \sqrt{2} \; \cos(2\, l\, \pi\, t),$$
$$\lambda_{2l} = \frac{2}{\pi^2 \; (2l)^2}, \quad \varphi_{2l}(t) = \sqrt{2} \; \sin(2\, l\, \pi\, t)$$

for $l \in I\!N$.

Theorem 2 *If* $P(Y_n - x(T_n)^\top \beta \geq 0 | T_n) = \frac{1}{2}$ *and* T_n *has continuous distribution, then*
a) the simplicial likelihood depth $d_S(\beta, (Z_1, \ldots, Z_N))$ *for linear regression satisfies*

$$\mathcal{L}\left(N \left(d_S(\beta, (Z_1, \ldots, Z_N)) - \frac{1}{4} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{Q}\left(\binom{3}{2} \lambda_l; \; l \in I\!N \right)$$

with $\lambda_{2l-1} = \frac{-1}{\pi^2\,(2l-1)^2}$ *and* $\lambda_{2l} = \frac{-1}{\pi^2\,(2l-1)^2}$ *for* $l \in I\!N$,
b) the simplicial likelihood depth $d_S(\beta, (Z_1, \ldots, Z_N))$ *for quadratic regression satisfies*

$$\mathcal{L}\left(N \left(d_S(\beta, (Z_1, \ldots, Z_N)) - \frac{1}{8} \right) \right) \xrightarrow{\mathcal{L}} \mathcal{Q}\left(\binom{4}{2} \lambda_l; \; l \in I\!N \cup \{0\} \right)$$

with $\lambda_0 = \frac{-1}{24}$, $\lambda_{2l-1} = \frac{-1}{\pi^2\,(2l)^2}$ *and* $\lambda_{2l} = \frac{-1}{\pi^2\,(2l)^2}$ *for* $l \in I\!N$.

There are several possibilities to calculate the quantiles of the distributions $\mathcal{Q}\left(\binom{3}{2}\lambda_l; \; l \in I\!N\right)$ and $\mathcal{Q}\left(\binom{4}{2}\lambda_l; \; l \in I\!N \cup \{0\}\right)$ in Theorem 2 (see e.g. Imhof (1961) or

α	5%	10%	15%	20%	25%	30%	35%
α-quantile	-1.038	-0.930	-0.839	-0.755	-0.674	-0.593	-0.510
99.5% bands	± 0.002	± 0.002	± 0.002	± 0.003	± 0.003	± 0.003	0.003
α	40%	45%	50%	55%	60%	65%	
α-quantile	-0.424	-0.334	-0.239	-0.136	-0.023	0.104	
99.5% bands	± 0.003	± 0.004	± 0.004	± 0.004	± 0.005	± 0.005	
α	70%	75%	80%	85%	90%	95%	
α-quantile	0.248	0.416	0.617	0.876	1.233	1.837	
99.5% bands	± 0.006	± 0.006	± 0.007	± 0.008	± 0.010	± 0.015	

Table 1: Means and 99.5% confidence bands of simulated quantiles for quadratic regression

Farebrother (1984)). One more simple possibility is the generation of random numbers of the distributions. For example, the quantiles for quadratic regression given in Table 1 were calculated by generating 10000 random numbers of the distribution $\mathcal{Q}\left(\binom{4}{2}\lambda_l;\ l \in \{0,\ldots,2L\}\right)$ for $L = 200$. The calculation of the quantiles was repeated 500 times. The means and standard errors (times $t(0.9975, 499)/\sqrt{500}$ where $t(\alpha, k)$ denotes the α-quantile of the t-distribution with k degrees of freedom) of these quantiles are given in Table 1. The same was done for $L = 100$. However, the results for $L = 100$ are very similar: The 99.5% confidence bands are even the same, only the means differ slightly in the last position.

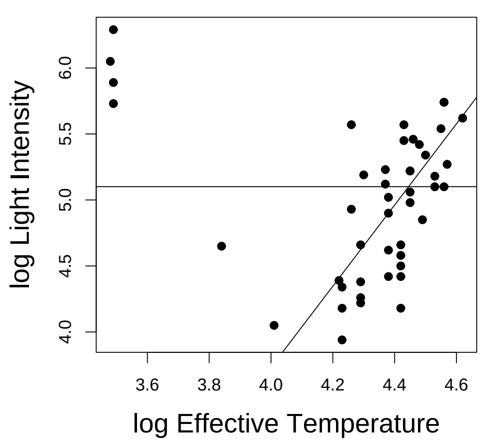

Figure 1: Hertzsprung-Russell data with catline and horizontal line through $\mu = 5.1$

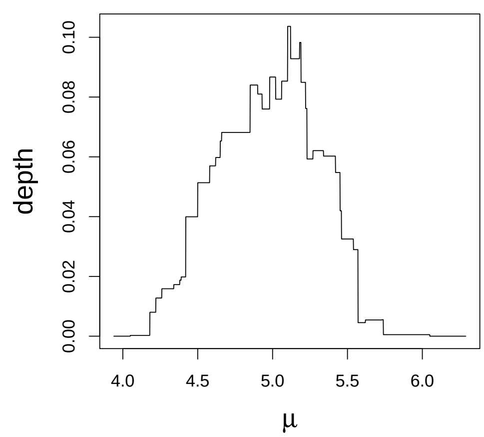

Figure 2: Depth of horizontal lines through μ for the Hertzsprung-Russell data

Example 7 (Hertzsprung-Russell data)
Figure 1 shows the Hertzsprung-Russell data introduced by Rousseeuw and Leroy (1987). These data concern the temperature and light intensities of 46 stars. Assuming a quadratic

regression model with parameter $\beta = (\mu, \beta_1, \beta_2)^\top$, we want to test the hypothesis that the true function is a constant function, i.e. $H_0 : \beta_1 = \beta_2 = 0$ or $H_0 : \beta \in \Theta_0$ where $\Theta_0 = \{\beta \in I\!R^3;\ \beta_1 = \beta_2 = 0\}$. The simplicial likelihood depth $d_S((\mu, 0, 0)^\top, (z_1, \ldots, z_N))$ for different horizontal lines through μ is plotted in Figure 2 by 2000 points between the minimum and maximum value of y_n. It turns out that $\sup_{\beta \in \Theta_0} d_S(\beta, (z_1, \ldots, z_N)) = \sup_{\mu \in I\!R} d_S((\mu, 0, 0)^\top, (z_1, \ldots, z_N)) = 0.104$ and that the maximum depth is attained by $\mu = 5.1$. Hence the test statistic according to (8) has the value -0.966 since $\gamma_\beta = (\frac{1}{2})^3$ and $N = 46$. Comparing this value with the 10%-quantile of Table 1 leads to a rejection of the hypothesis for the significance level 10%.

For testing the hypothesis that the regression function is linear, i.e. $H_0 : \beta_2 = 0$ or $H_0 : \beta \in \Theta_0$ where $\Theta_0 = \{\beta \in I\!R^3;\ \beta_2 = 0\}$, the catline of Hubert and Rousseeuw (1998) was calculated and plotted in Figure 1. It has the parameter $(\mu, \beta_1, \beta_2) = (-8.6, 3.1, 0)$ and its simplicial likelihood depth is 0.134. Hence the test statistic according to (8) satisfies $\sup_{\beta \in \Theta_0} N(d_S(\beta, (z_1, \ldots, z_N)) - \gamma_\beta) \geq 46 * (0.134 - 0.125) = 0.414$ which is larger than the 70%-quantile of Table 1. Hence the hypothesis can not be rejected.

Note that the classical F-test provides for $H_0 : \beta_1 = \beta_2 = 0$ and $H_0 : \beta_2 = 0$ a p-value less than 0.0001. This is due to the outliers, giants, in the left upper corner of Figure 1. Without these outliers, a linear regression line is a good description of the data. Hence the test for $H_0 : \beta_2 = 0$ based on the simplicial likelihood depth is outlier robust. However, a horizontal line is not a good description of the data. But the test based on the simplicial likelihood depth rejects this hypothesis only with respect to the significance level 10%. Hence the efficiency of this test is not so good as for the classical test. But this is the case for all nonparametric tests. □

5 Conclusion and open problems

The possibility to base tests on the simplicial likelihood depth is a tractable way of deriving tests for polynomial regression. Although it is only demonstrated up to quadratic regression it seems reasonable that this can be done with the same method also for polynomial regression of higher order. There, differential equations of higher order appear so that the set of possible solutions is larger which make the calculations longer and more tedious.

An open problem is the calculation of $\sup_{\theta \in \Theta_0} d_S(\theta, (z_1, \ldots, z_N))$. A simple possibility is to use a global search based on all polynomials of the hypothesis through q points like in Example 7. Certainly there are better methods similar to those proposed for maximum regression depth estimators by Rousseeuw and Hubert (1999), Rousseeuw and Struyf (1998), Van Aelst et al. (2002). An open problem is also the question whether the presented method can be used for other problems like multiple regression and quite different models. While likelihood depth and likelihood depth estimators for regression with observations with discrete distributions can be derived via the method for regression depth of Rousseeuw and Hubert (1999), the proposed method for deriving tests is not working for discrete distributions of observations. The tests can be based on the simplicial likelihood depth but $E(\psi_\theta(Z_1,\ldots,Z_{q+1})|Z_1 = z_1)$ cannot be derived as presented since $P(\tilde{Y}_n - x(T_n)^\top\beta \geq 0|T_n)$ is not constant even if $\tilde{Y}_n$ is the appropriate transformed observation. Hence alternative methods for calculating $E(\psi_\theta(Z_1,\ldots,Z_{q+1})|Z_1 = z_1)$ also for polynomial regression must be found. It is very likely that $E(\psi_\theta(Z_1,\ldots,Z_{q+1})|Z_1 = z_1)$ is not independent of z_1 as for exponential distribution so that the simplicial likelihood depth would not be a degenerated U-statistic.

6 Proofs

Lemma 3 *If $T_1,\ldots,T_{n+1}$ are i.i.d. with differentiable distribution function G and $t,s \in \mathbb{R}$, then*

$$a)\quad P(T_1 < T_2 < \ldots < T_{n+1}|T_1 = t) = \sum_{i=0}^{n}(-1)^i \frac{1}{(n-i)!\; i!}\, G(t)^i,$$

$$b)\quad P(T_1 < T_2 < \ldots < T_{n+1}|T_{n+1} = t) = \frac{1}{n!}\, G(t)^n,$$

$$c)\quad P(T_1 < T_2 < \ldots < T_{m+1} < \ldots < T_{n+1}|T_{m+1} = t) = \frac{1}{m!}\sum_{i=0}^{n-m}(-1)^i \frac{1}{(n-m-i)!\; i!}\, G(t)^{m+i} \quad \text{for } m = 0,1,\ldots,n,$$

$$d)\quad \sum_{m=0}^{n} P(T_1 < T_2 < \ldots < T_{m+1} < \ldots < T_{n+1}|T_{m+1} = t) = \frac{1}{n!}$$

$$e)\quad P(t < T_1 < \ldots < T_n < s) = \sum_{i=0}^{n}(-1)^i \frac{1}{(n-i)!\; i!}\, G(s)^{n-i}\, G(t)^i.$$

Proof of Lemma 3.
Using $\int_a^b G(x)^k\, g(x)\, dx = \frac{1}{k+1}\left(G(b)^{k+1} - G(a)^{k+1}\right)$ for $g = G'$, the assertions a) und b) can be proved by induction over n. The assertion of c) is obtained by using a) and b) since independence implies

$$\begin{aligned} &P(T_1 < T_2 < \ldots < T_{m+1} < \ldots < T_{n+1}|T_{m+1} = t) \\ &= P(T_1 < T_2 < \ldots < T_m < t)\; P(t < T_{m+2} < \ldots < T_{n+1}). \end{aligned}$$

By summing over the probabilities of c), the assertion d) follows. Induction over n provides also the assertion e).□

Lemma 4 *Let $T_1, \dots, T_n$ be i.i.d. with differentiable distribution function G, $t, s \in I\!R$ with $t < s$, and define for $k, l, m \in I\!N \cup \{0\}$ with $k + l + m = n$*

$$\alpha(k+1, k+l+2) := \sum_{\pi \in \mathcal{P}_n} P\big(T_{\pi(1)} < \dots < T_{\pi(k)} < t < T_{\pi(k+1)} \\ < \dots < T_{\pi(k+l)} < s < T_{\pi(k+l+1)} < \dots < T_{\pi(n)}\big).$$

Then

$$a) \quad \alpha(k+1, k+l+2) = \frac{(k+l+m)!}{k!\; l!\; m!}\, G(t)^k\, (G(s) - G(t))^l\, (1 - G(s))^m,$$

$$b) \quad \sum_{k=0}^{n} \sum_{l=0}^{n-k} \alpha(k+1, k+l+2) = 1,$$

$$c) \quad \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{n-2l} \alpha(k+1, k+2l+2) = \frac{1}{2}\, (1 - 2|G(s) - G(t)|)^n + \frac{1}{2}.$$

Proof of Lemma 4.
a) The independence assumption and Lemma 3 a), b) and e) imply

$$\begin{aligned} & P(T_1 < \dots < T_k < t < T_{k+1} < \dots < T_{k+l} < s < T_{k+l+1} < \dots < T_n) \\ = \;& P(T_1 < \dots < T_k < t)\; P(t < T_{k+1} < \dots < T_{k+l} < s)\; P(s < T_{k+l+1} < \dots < T_n) \\ = \;& \frac{G(t)^k}{k!\; l!\; m!} \sum_{i=0}^{l} \binom{l}{i} G(s)^{l-i}\, G(t)^i\, (-1)^i \sum_{j=0}^{m} \binom{m}{j} (-1)^j\, G(s)^j \\ = \;& \frac{1}{k!\; l!\; m!}\, G(t)^k\, (G(s) - G(t))^l\, (1 - G(s))^m. \end{aligned}$$

b) Part a) imply

$$\begin{aligned} & \sum_{k}^{n} \sum_{l=0}^{n-k} \alpha(k+1, k+l+2) \\ = \;& \sum_{k}^{n} \frac{(k+l+n-k-l)!}{k!\; (n-k)!}\, G(t)^k \sum_{l=0}^{n-k} \frac{(n-k)!}{l!\; (n-k-l)!}\, (G(s) - G(t))^l\, (1 - G(s))^{n-k-l} \\ = \;& \sum_{k}^{n} \frac{n!}{k!\; (n-k)!}\, G(t)^k\, (G(s) - G(t) + 1 - G(s))^{n-k} \\ = \;& (G(t) + 1 - G(t))^n = 1. \end{aligned}$$

c) Part a) imply similarly as in b)

$$\begin{aligned}
&\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{n-2l} \alpha(k+1, k+2l+2) \\
&= \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0}^{n-2l} \binom{n-2l}{k} (-1)^k (G(s)-G(t))^k \\
&= \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0,\ k \text{ even}}^{n-2l} \binom{n-2l}{k} (G(s)-G(t))^k \\
&\quad + \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0,\ k \text{ odd}}^{n-2l} \binom{n-2l}{k} (-1)^k (G(s)-G(t))^k.
\end{aligned}$$

For even k we have

$$\begin{aligned}
&\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0,\ k \text{ even}}^{n-2l} \binom{n-2l}{k} (G(s)-G(t))^k \\
&= \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0}^{\lfloor \frac{n}{2}-l \rfloor} \binom{n-2l}{2k} (G(s)-G(t))^{2k} \\
&= \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{k=l}^{\lfloor \frac{n}{2} \rfloor} (G(s)-G(t))^{2k} \binom{n}{2l} \binom{n-2l}{2k-2l} \\
&= \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (G(s)-G(t))^{2k} \binom{n}{2k} \sum_{l=0}^{k} \binom{2k}{2l} \\
&= \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} (G(s)-G(t))^{2k} \binom{n}{2k} \sum_{l=0}^{k} \binom{2k}{2l} + 1.
\end{aligned}$$

For odd k we obtain similarly

$$\begin{aligned}
&\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=0,\ k \text{ odd}}^{n-2l} \binom{n-2l}{k} (-1)^k (G(s)-G(t))^k \\
&= \sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \binom{n}{2l} (G(s)-G(t))^{2l} \sum_{k=1}^{\lfloor \frac{n+1}{2}-l \rfloor} \binom{n-2l}{2k-1} (-1) (G(s)-G(t))^{2k-1} \\
&= -\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{k=l+1}^{\lfloor \frac{n+1}{2} \rfloor} (G(s)-G(t))^{2k-1} \binom{n}{2l} \binom{n-2l}{2k-2l-1} \\
&= -\sum_{k=1}^{\lfloor \frac{n+1}{2} \rfloor} (G(s)-G(t))^{2k-1} \binom{n}{2k-1} \sum_{l=0}^{k-1} \binom{2k-1}{2l}.
\end{aligned}$$

Moreover we have

$$\sum_{l=0}^{k} \binom{2k}{2l} = 2^{2k-1} \quad \text{and} \quad \sum_{l=0}^{k-1} \binom{2k-1}{2l} = 2^{2k-2}, \tag{12}$$

where the second equality in (12) can be seen by the equality

$$2 \sum_{l=0}^{k-1} \binom{2k-1}{2l} = \sum_{l=0}^{2k-1} \binom{2k-1}{l}.$$

For showing the first equality in (12), induction over k and the property $\binom{n}{k} + \binom{n}{k+2} = \binom{n+2}{k+1} - 2\binom{n}{k+1}$ are needed additionally.

Hence we obtain

$$\begin{aligned}
&\sum_{l=0}^{\lfloor \frac{n}{2} \rfloor} \sum_{k=0}^{n-2l} \alpha(k+1, k+2l+2) \\
&= \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} (G(s) - G(t))^{2k} \binom{n}{2k} 2^{2k-1} + 1 - \sum_{k=1}^{\lfloor \frac{n+1}{2} \rfloor} (G(s) - G(t))^{2k-1} \binom{n}{2k-1} 2^{2k-2} \\
&= 1 + \frac{1}{2} \sum_{k=1}^{n} |G(s) - G(t)|^k \, 2^k \, (-1)^k \binom{n}{k} \\
&= \frac{1}{2} \, (1 - 2|G(s) - G(t)|)^n + \frac{1}{2}. \square
\end{aligned}$$

Proof of Proposition 1.
Let

$$\mathcal{P}_{q+1}(m) := \{\pi : \{1, \ldots, q+1\} \setminus \{m\} \to \{1, \ldots, q+1\} \setminus \{m\};\ \pi(i) \neq \pi(j) \text{ for } i \neq j\}$$

the set of all permutations of $\{1, \ldots, q+1\} \setminus \{m\}$ and set $R_n := Y_n - x(T_n)^\top \beta$, $r_n := y_n - x(t_n)^\top \beta$ for the residuals.
a) Since the indicator variables $1\{R_n \geq 0\}$ and the explanatory variables T_n are indepen-

dent, we have

$$
\begin{aligned}
& E\left(\psi_\beta((Z_1,\ldots,Z_{q+1})|Z_1=(y_1,t_1))\right) \\
= & \sum_{\pi\in\mathcal{P}_{q+1}} \left(\prod_{i=1}^{q+1} P\left(R_{\pi(i)}\left(-1\right)^i \geq 0|R_1=r_1\right) + \prod_{i=1}^{q+1} P\left(R_{\pi(i)}\left(-1\right)^i \leq 0|R_1=r_1\right) \right) \\
& \cdot P(T_{\pi(1)} < T_{\pi(2)} < \ldots < T_{\pi(q+1)}|T_1=t_1) \\
= & \sum_{m=1}^{q+1} \sum_{\pi\in\mathcal{P}_{q+1}(m)} \Big(p^{\frac{q+1}{2}-1}\left(1-p\right)^{\frac{q+1}{2}} P\left(R_m \geq 0|R_m=r_m\right) \\
& \qquad + p^{\frac{q+1}{2}}\left(1-p\right)^{\frac{q+1}{2}-1} P\left(R_m \leq 0|R_m=r_m\right)\Big) \\
& \cdot P(T_{\pi(1)} < T_{\pi(2)} < \ldots < T_{\pi(m-1)} < T_m < T_{\pi(m+1)} < \ldots < T_{\pi(q+1)}|T_m=t_1) \\
= & \sum_{m=1}^{q+1} q!\, p^{\frac{q+1}{2}-1}\left(1-p\right)^{\frac{q+1}{2}-1}\left(\left(1-p\right)1\{r_m \geq 0\} + p\,1\{r_m \leq 0\}\right) \\
& \cdot P(T_1 < T_2 < \ldots < T_{q+1}|T_m=t_1) \\
= & p^{\frac{q+1}{2}-1}\left(1-p\right)^{\frac{q+1}{2}-1} q! \\
& \cdot \Big(\left(1-p\right)1\{y_1 - x(t_1)^\top\beta \geq 0\} + p\,1\{y_1 - x(t_1)^\top\beta \leq 0\}\Big) \\
& \qquad \cdot \sum_{m=1}^{q+1} P(T_1 < T_2 < \ldots < T_{q+1}|T_m=t_1).
\end{aligned}
$$

Lemma 3 d) provides then the assertion.
b) Analogously to a) we have

$$
\begin{aligned}
& E\left(\psi_\beta((Z_1,\ldots,Z_{q+1})|Z_1=(y_1,t_1))\right) \\
= & \sum_{m=0}^{\frac{q}{2}} \sum_{\pi\in\mathcal{P}_{q+1}(2m+1)} \Big(p^{\frac{q}{2}}\left(1-p\right)^{\frac{q}{2}} P\left(R_{2m+1} \geq 0|R_{2m+1}=r_{2m+1}\right) \\
& \qquad + p^{\frac{q}{2}}\left(1-p\right)^{\frac{q}{2}} P\left(R_{2m+1} \leq 0|R_{2m+1}=r_{2m+1}\right)\Big) \\
& \cdot P(T_{\pi(1)} < T_{\pi(2)} < \ldots < T_{\pi(2m)} < T_{2m+1} < T_{\pi(2m+2)} < \ldots < T_{\pi(q+1)}|T_{2m+1}=t_1) \\
& + \sum_{m=1}^{\frac{q}{2}} \sum_{\pi\in\mathcal{P}_{q+1}(2m)} \Big(p^{\frac{q}{2}-1}\left(1-p\right)^{\frac{q}{2}+1} P\left(R_{2m} \geq 0|R_{2m}=r_{2m}\right) \\
& \qquad + p^{\frac{q}{2}+1}\left(1-p\right)^{\frac{q}{2}-1} P\left(R_{2m} \leq 0|R_{2m}=r_{2m}\right)\Big) \\
& \cdot P(T_{\pi(1)} < T_{\pi(2)} < \ldots < T_{\pi(2m-1)} < T_{2m} < T_{\pi(2m+1)} < \ldots < T_{\pi(q+1)}|T_{2m}=t_1). \square
\end{aligned}
$$

Proof of Proposition 2.
Set

$$\alpha(k+1,k+l+2,t_1,t_2) := \sum_{\pi\in\mathcal{P}_{q-1}} P\big(T_{\pi(1)} < \ldots < T_{\pi(k)} < t_1 < T_{\pi(k+1)} < \ \ldots < T_{\pi(k+l)} < t_2 < T_{\pi(k+l+1)} < \ldots < T_{\pi(q-1)}\big)$$

for $t_1 < t_2$ and

$$H \ := \ \frac{1}{2}\ (1-2|G(t_1)-G(t_2)|)^{q-1} + \frac{1}{2}.$$

According to Lemma 4 b) and c) we have

$$\begin{aligned} H &= \sum_{l=0}^{\lfloor\frac{q-1}{2}\rfloor} \sum_{k=0}^{q-1-2l} (\alpha(k+1,k+2l+2,t_1,t_2)\ 1\{t_1<t_2\} \\ &\quad + \ \alpha(k+1,k+2l+2,t_2,t_1)\ 1\{t_2<t_1\}) \\ &= \sum_{k=0}^{q-1} \sum_{l=0,\ l \text{ even}}^{q-1-k} (\alpha(k+1,k+l+2,t_1,t_2)\ 1\{t_1<t_2\} \\ &\quad + \ \alpha(k+1,k+l+2,t_2,t_1)\ 1\{t_2<t_1\}) \end{aligned}$$

and

$$\begin{aligned} &\sum_{k=0}^{q-1} \sum_{l=0,\ l \text{ odd}}^{q-1-k} (\alpha(k+1,k+l+2,t_1,t_2)\ 1\{t_1<t_2\} \\ &\quad + \ \alpha(k+1,k+l+2,t_2,t_1)\ 1\{t_2<t_1\}) \\ &= \left(1-\sum_{k=0}^{q-1} \sum_{l=0,\ l \text{ even}}^{q-1-k} \alpha(k+1,k+l+2,t_1,t_2)\right)\ 1\{t_1<t_2\} \\ &\quad + \left(1-\sum_{k=0}^{q-1} \sum_{l=0,\ l \text{ even}}^{q-1-k} \alpha(k+1,k+l+2,t_2,t_1)\right)\ 1\{t_2<t_1\} \\ &= \ 1-H. \end{aligned}$$

This implies because of the independence of the residuals $R_n = Y_n - x(T_n)^\top\beta$ and T_n

$$
\begin{aligned}
&E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,Z_2=z_2)\\
&= \sum_{\pi\in\mathcal{P}_{q+1}} \left(P\left(\bigcap_{k=1}^{q+1} \{R_{\pi(k)}\ (-1)^k \geq 0\} | R_1=r_1, R_2=r_2 \right) \right.\\
&\qquad \left. +\ P\left(\bigcap_{k=1}^{q+1} \{R_{\pi(k)}\ (-1)^k \leq 0\} | R_1=r_1, R_2=r_2 \right) \right)\\
&\quad \cdot P(T_{\pi(1)} < \ldots < T_{\pi(q+1)} | T_1=t_1, T_2=t_2)\\
&= \sum_{\substack{\pi\in\mathcal{P}_{q+1}\\ \pi^{-1}(1)-\pi^{-1}(2)\ \text{odd}}} \left(\frac{1}{2}\right)^{q-1} (1\{r_1\geq 0\}\ 1\{r_2\leq 0\} + 1\{r_1\leq 0\}\ 1\{r_2\geq 0\})\\
&\qquad\qquad \cdot\ P(T_{\pi(1)} < \ldots < T_{\pi(q+1)} | T_1=t_1, T_2=t_2)\\
&\quad + \sum_{\substack{\pi\in\mathcal{P}_{q+1}\\ \pi^{-1}(1)-\pi^{-1}(2)\ \text{even}}} \left(\frac{1}{2}\right)^{q-1} (1\{r_1\geq 0\}\ 1\{r_2\geq 0\} + 1\{r_1\leq 0\}\ 1\{r_2\leq 0\})\\
&\qquad\qquad \cdot\ P(T_{\pi(1)} < \ldots < T_{\pi(q+1)} | T_1=t_1, T_2=t_2)\\
&= \left(\frac{1}{2}\right)^{q-1} \tau(r_1,r_2) \sum_{\substack{\pi\in\mathcal{P}_{q+1}\\ \pi^{-1}(1)-\pi^{-1}(2)\ \text{odd}}} P(T_{\pi(1)} < \ldots < T_{\pi(q+1)} | T_1=t_1, T_2=t_2)\\
&\quad + \left(\frac{1}{2}\right)^{q-1} (1-\tau(r_1,r_2)) \sum_{\substack{\pi\in\mathcal{P}_{q+1}\\ \pi^{-1}(1)-\pi^{-1}(2)\ \text{even}}} P(T_{\pi(1)} < \ldots < T_{\pi(q+1)} | T_1=t_1, T_2=t_2)\\
&= \left(\frac{1}{2}\right)^{q-1} \tau(r_1,r_2) \sum_{k=0}^{q-1} \sum_{l=0,\ l\ \text{even}}^{q-1-k} (\alpha(k+1,k+l+2,t_1,t_2)\ 1\{t_1<t_2\}\\
&\qquad +\ \alpha(k+1,k+l+2,t_2,t_1)\ 1\{t_2<t_1\})\\
&\quad + \left(\frac{1}{2}\right)^{q-1} (1-\tau(r_1,r_2)) \sum_{k=0}^{q-1} \sum_{l=0,\ l\ \text{odd}}^{q-1-k} (\alpha(k+1,k+l+2,t_1,t_2)\ 1\{t_1<t_2\}\\
&\qquad +\ \alpha(k+1,k+l+2,t_2,t_1)\ 1\{t_2<t_1\})\\
&= \left(\frac{1}{2}\right)^{q-1} \tau(r_1,r_2)\ H + \left(\frac{1}{2}\right)^{q-1} (1-\tau(r_1,r_2))\ (1-H)\\
&= \left(\frac{1}{2}\right)^{q-1} \left(\left(\tau(r_1,r_2) - \frac{1}{2} \right) (2H-1) + \frac{1}{2} \right)\\
&= \left(\frac{1}{2}\right)^{q-1} \left(\tau(r_1,r_2) - \frac{1}{2} \right) (1-2|G(t_1)-G(t_2)|)^{q-1} + \left(\frac{1}{2}\right)^{q} . \Box
\end{aligned}
$$

Proof of Lemma 1
Since

$$\begin{aligned}
&\int_0^1 \left(\frac{1}{2} - |t-s|\right) \varphi(t)\, dt \\
&= \frac{1}{2}\int_0^1 \varphi(t)\, dt - \int_0^s (s-t)\,\varphi(t)\, dt - \int_s^1 (t-s)\,\varphi(t)\, dt \\
&= \frac{1}{2}\int_0^1 \varphi(t)\, dt - 2\,s \int_0^s \varphi(t)\, dt + 2 \int_0^s t\,\varphi(t)\, dt \\
&\quad - \int_0^1 t\,\varphi(t)\, dt + s \int_0^1 \varphi(t)\, dt,
\end{aligned} \tag{13}$$

the differentiation of $\lambda\,\varphi(s) = \int_0^1 \left(\frac{1}{2} - |t-s|\right)\ \varphi(t)\, dt$ two times leads to

$$\begin{aligned}
&\lambda\,\varphi'(s) = -2 \int_0^s \varphi(t)\, dt - 2\,s\,\varphi(s) + 2\,s\,\varphi(s) + \int_0^1 \varphi(t)\, dt, \\
&\lambda\,\varphi''(s) = -2\,\varphi(s) \quad \text{or} \quad \varphi''(s) + \frac{2}{\lambda}\,\varphi(s) = 0.
\end{aligned}$$

The solutions of the last differential equation have the form (see e.g. Kamke 1947, p. 252, or Brauer and Nohel 1968, p. 88)

$$\varphi(s) = c_1\, exp\left(s\,\sqrt{-\frac{2}{\lambda}}\right) + c_2\, exp\left(-s\,\sqrt{-\frac{2}{\lambda}}\right), \tag{14}$$

if $\frac{2}{\lambda} < 0$, and

$$\varphi(s) = c_1\,\cos\left(s\,\sqrt{\frac{2}{\lambda}}\right) + c_2\,\sin\left(s\,\sqrt{\frac{2}{\lambda}}\right), \tag{15}$$

if $\frac{2}{\lambda} > 0$.

Now set $\psi(s) := \sqrt{\frac{|\lambda|}{2}}\ \varphi'(s)$. Then we have $\varphi'(s) = \sqrt{\frac{2}{|\lambda|}}\ \psi(s)$ and $-\frac{2}{\lambda}\,\varphi(s) = \varphi''(s) = \sqrt{\frac{2}{|\lambda|}}\ \psi'(s)$ so that $-\text{sgn}(\lambda)\ \sqrt{\frac{|\lambda|}{2}}\ \psi'(s) = \varphi(s)$ and

$$\int_0^s \varphi(t)\, dt = -\text{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\ (\psi(s) - \psi(0)), \tag{16}$$

$$\begin{aligned}
\int_0^s t\,\varphi(t)\, dt &= -\text{sgn}(\lambda)\ \sqrt{\frac{|\lambda|}{2}}\ t\,\psi(t)\Big|_0^s + \text{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}} \int_0^s \psi(t)\, dt \\
&= -\text{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\ s\,\psi(s) + \text{sgn}(\lambda)\,\frac{|\lambda|}{2}\ (\varphi(s) - \varphi(0)).
\end{aligned} \tag{17}$$

Using these properties in (13), we obtain

$$\begin{aligned}
&\lambda\,\varphi(s)\\
=\;& -\frac{1}{2}\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) + 2\,s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(s)-\psi(0))\\
& - 2\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,s\,\psi(s) + 2\,\mathrm{sgn}(\lambda)\,\frac{|\lambda|}{2}\,(\varphi(s)-\varphi(0))\\
& + \mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\psi(1) - \mathrm{sgn}(\lambda)\,\frac{|\lambda|}{2}\,(\varphi(1)-\varphi(0))\\
& - s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0))\\
=\;& \lambda\,\varphi(s) + s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\left(-2\,\psi(0)-\psi(1)+\psi(0)\right)\\
& + \mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\left(-\frac{1}{2}\,\psi(1)+\frac{1}{2}\,\psi(0)+\psi(1)\right) + \frac{\lambda}{2}\,(-2\varphi(0)-\varphi(1)+\varphi(0))\,.
\end{aligned}$$

This implies $\psi(0)+\psi(1)=0$ and $\varphi(0)+\varphi(1)=0$ or

$$\varphi(0)+\varphi(1)=0 \quad \text{and} \quad \varphi'(0)+\varphi'(1)=0. \tag{18}$$

Now set $a := \sqrt{\frac{2}{|\lambda|}}$. If $\lambda<0$, then any eigenfunction φ must satisfy (14) and (18). This means

$$\begin{aligned}
0 \;&=\; \varphi(0)+\varphi(1) = c_1(1+\exp(a)) + c_2(1+\exp(-a))\\
&\iff c_2 = -c_1\,\frac{1+\exp(a)}{1+\frac{1}{\exp(a)}} = -c_1\;\exp(a),\\
0 \;&=\; \varphi'(0)+\varphi'(1) = a\;c_1(1+\exp(a)) - a\;c_2(1+\exp(-a))\\
&\iff c_2 = c_1\,\frac{1+\exp(a)}{1+\frac{1}{\exp(a)}} = c_1\;\exp(a),
\end{aligned}$$

implying $c_1=0=c_2$. Hence the eigenfunction cannot have the form (14).

If $\lambda>0$, then the eigenfunction φ must satisfy (15) and (18). Then we obtain

$$0 \;=\; \varphi(0)+\varphi(1) = c_1(1+\cos(a)) + c_2\;\sin(a), \tag{19}$$
$$0 \;=\; \varphi'(0)+\varphi'(1) = a\;\left(c_2(1+\cos(a)) - c_1\;\sin(a)\right). \tag{20}$$

Both equations (19) and (20) are satisfied for $a=(2k+1)\pi$ with $k\in\mathbb{Z}$. For $a\neq(2k+1)\pi$, equation (19) implies

$$c_1 = -c_2\,\frac{\sin(a)}{1+\cos(a)}.$$

Plugging this in (20) yields

$$\begin{aligned} 0 &= c_2 \left(1+\cos(a)+\frac{\sin(a)^2}{1+\cos(a)}\right) \\ &\iff 0 = 1+2\cos(a)+\cos(a)^2+\sin(a)^2 = 2+2\cos(a), \end{aligned}$$

which implies $-1 = \cos(a)$ and thus the contradiction $a = (2k+1)\pi$. Hence the eigenfunctions can be only of the form

$$\varphi(s) = c_1\ \cos(as) + c_2\ \sin(as)$$

with $a = (2k+1)\,\pi$ and $k \in \mathbb{Z}$. Since $\cos((2k+1)\pi\ s) = \cos(-(2k+1)\pi\ s)$ and $\sin((2k+1)\pi\ s) = -\sin(-(2k+1)\pi\ s)$ we can restrict ourselves to $k \in \mathbb{N} \cup \{0\}$. Under this restriction the functions

$$\left\{\sqrt{2}\ \cos((2k-1)\pi\ s);\ k \in \mathbb{N}\right\} \cup \left\{\sqrt{2}\ \sin((2k-1)\pi\ s);\ k \in \mathbb{N}\right\}$$

are orthogonal and normalized. Since $a^2 := \frac{2}{|\lambda|} = \frac{2}{\lambda}$, the corresponding eigenvalues are given by $\lambda = \frac{2}{(2k-1)^2\ \pi^2}$ with $k \in \mathbb{N}$. □

Proof of Lemma 2

Using similar arguments as in (13) we have

$$\begin{aligned} &\int_0^1 \left(\frac{1}{4} - |t-s| + (t-s)^2\right)\ \varphi(t)\ dt \\ &= \frac{1}{4}\int_0^1 \varphi(t)\ dt - 2\,s \int_0^s \varphi(t)\ dt + 2\int_0^s t\ \varphi(t)\ dt \\ &\quad - \int_0^1 t\ \varphi(t)\ dt + s\int_0^1 \varphi(t)\ dt \\ &\quad + \int_0^1 t^2\ \varphi(t)\ dt - 2\,s \int_0^1 t\ \varphi(t)\ dt + s^2 \int_0^1 \varphi(t)\ dt. \end{aligned} \tag{21}$$

Differentiation of $\lambda\,\varphi(s) = \int_0^1 \left(\frac{1}{4} - |t-s| + (t-s)^2\right)\ \varphi(t)\ dt$ three times leads to

$$\begin{aligned} \lambda\,\varphi'(s) &= -2\int_0^s \varphi(t)\ dt - 2\,s\,\varphi(s) + 2\,s\,\varphi(s) + \int_0^1 \varphi(t)\ dt \\ &\quad - 2\int_0^1 t\ \varphi(t)\ dt + 2\,s\int_0^1 \varphi(t)\ dt, \\ \lambda\,\varphi''(s) &= -2\,\varphi(s) + 2\int_0^1 \varphi(t)\ dt, \\ \lambda\,\varphi'''(s) &= -2\,\varphi'(s) \quad \text{or} \quad \varphi'''(s) + \frac{2}{\lambda}\,\varphi'(s) = 0. \end{aligned}$$

The solutions of the last differential equation have the form (see e.g. Kamke 1947, p. 252, or Brauer and Nohel 1968, p. 88)

$$\varphi(s) = c_1\, exp\left(s\, \sqrt{-\frac{2}{\lambda}} \right) + c_2\, exp\left(-s\, \sqrt{-\frac{2}{\lambda}} \right) + c_3, \tag{22}$$

if $\frac{2}{\lambda} < 0$, and

$$\varphi(s) = c_1\, \cos\left(s\, \sqrt{\frac{2}{\lambda}} \right) + c_2\, \sin\left(s\, \sqrt{\frac{2}{\lambda}} \right) + c_3, \tag{23}$$

if $\frac{2}{\lambda} > 0$.

In both cases (22) and (23), the solution φ can be written as $\varphi(s) = \tilde{\varphi}(s) + c_3$. Now set $\psi(s) := \sqrt{\frac{|\lambda|}{2}}\, \tilde{\varphi}'(s)$. Then we have in both cases $-\text{sgn}(\lambda)\, \sqrt{\frac{|\lambda|}{2}}\, \psi'(s) = \tilde{\varphi}(s)$, i.e. the same property which φ satisfied in the proof of Lemma 1. Besides the properties (16) and (17) used in the proof of Lemma 1, we will use

$$\begin{aligned}
&\int_0^s t^2\, \tilde{\varphi}(t)\, dt = -\text{sgn}(\lambda)\, \sqrt{\frac{|\lambda|}{2}}\, t^2\, \psi(t)\Big|_0^s + \text{sgn}(\lambda)\, 2\, \sqrt{\frac{|\lambda|}{2}} \int_0^s t\, \psi(t)\, dt \\
&= -\text{sgn}(\lambda)\, \sqrt{\frac{|\lambda|}{2}}\, s^2\, \psi(s) + \text{sgn}(\lambda)|\lambda|\, t\, \tilde{\varphi}(t)|_0^s - \text{sgn}(\lambda)|\lambda| \int_0^s \tilde{\varphi}(t)\, dt \\
&= -\text{sgn}(\lambda)\, \sqrt{\frac{|\lambda|}{2}}\, s^2\, \psi(s) + \lambda\, s\, \tilde{\varphi}(s) + |\lambda|\, \sqrt{\frac{|\lambda|}{2}}\, (\psi(s) - \psi(0))
\end{aligned} \tag{24}$$

and

$$\begin{aligned}
&\int_0^1 \left(\frac{1}{4} - |t-s| + (t-s)^2 \right)\, dt \\
&= \frac{1}{4} - \int_0^s (s-t)\, dt - \int_s^1 (t-s)\, dt + \int_0^1 (t^2 - 2st + s^2)\, dt \\
&= \frac{1}{4} - s^2 + \frac{1}{2}\, s^2 - \frac{1}{2} + \frac{1}{2}\, s^2 + s(1-s) + \frac{1}{3} - s + s^2 \\
&= \frac{1}{12}.
\end{aligned} \tag{25}$$

Plugging the properties (16), (17), (24), (25) in (21), we obtain

$$
\begin{aligned}
\lambda\,\varphi(s) &= \lambda\,\tilde{\varphi}(s) + \lambda\, c_3 \\
&= -\frac{1}{4}\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) + 2\,s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(s)-\psi(0)) \\
&\quad - 2\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,s\,\psi(s) + 2\,\mathrm{sgn}(\lambda)\,\frac{|\lambda|}{2}\,(\tilde{\varphi}(s)-\tilde{\varphi}(0)) \\
&\quad + \mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\psi(1) - \mathrm{sgn}(\lambda)\,\frac{|\lambda|}{2}\,(\tilde{\varphi}(1)-\tilde{\varphi}(0)) \\
&\quad - s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) \\
&\quad - \mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\psi(1) + \lambda\,\tilde{\varphi}(1) + |\lambda|\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) \\
&\quad + 2\,s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,\psi(1) - s\,\lambda\,(\tilde{\varphi}(1)-\tilde{\varphi}(0)) \\
&\quad - s^2\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) \\
&\quad + c_3\,\frac{1}{12} \\
&= \lambda\,\tilde{\varphi}(s) - \; s^2\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) \\
&\quad + s\,\mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\;(-2\,\psi(0)-\psi(1)+\psi(0)+2\psi(1)) \\
&\quad -s\,\lambda\,(\tilde{\varphi}(1)-\tilde{\varphi}(0)) \\
&\quad + \mathrm{sgn}(\lambda)\,\sqrt{\frac{|\lambda|}{2}}\;\left(-\frac{1}{4}\,\psi(1)+\frac{1}{4}\,\psi(0)+\psi(1)-\psi(1)\right) \\
&\quad + \frac{\lambda}{2}\,(-2\tilde{\varphi}(0)-\tilde{\varphi}(1)+\tilde{\varphi}(0)+2\tilde{\varphi}(1)) \\
&\quad + |\lambda|\,\sqrt{\frac{|\lambda|}{2}}\,(\psi(1)-\psi(0)) \\
&\quad + c_3\,\frac{1}{12}.
\end{aligned}
$$

This implies $\psi(1)-\psi(0)=0$ and $\tilde{\varphi}(1)-\tilde{\varphi}(0)=0$ or

$$\tilde{\varphi}(1)-\tilde{\varphi}(0)=0 \quad \text{and} \quad \tilde{\varphi}'(1)-\tilde{\varphi}'(0)=0. \tag{26}$$

We also get $\lambda=\frac{1}{12}$ or $c_3=0$. Now set $a := \sqrt{\frac{2}{|\lambda|}}$.

If $\lambda < 0$, then any eigenfunction φ must satisfy (22) and (26). This means

$$\begin{aligned} 0 &= \tilde{\varphi}(1) - \tilde{\varphi}(0) = c_1(\exp(a) - 1) + c_2(\exp(-a) - 1) \\ &\iff c_2 = -c_1 \, \frac{\exp(a) - 1}{\frac{1}{\exp(a)} - 1} = c_1 \, \exp(a), \\ 0 &= \tilde{\varphi}'(1) - \tilde{\varphi}'(0) = a\, c_1(\exp(a) - 1) - a\, c_2(\exp(-a) - 1) \\ &\iff c_2 = c_1 \, \frac{\exp(a) - 1}{\frac{1}{\exp(a)} - 1} = -c_1 \, \exp(a), \end{aligned}$$

implying $c_1 = 0 = c_2$. Since $\lambda = \frac{1}{12} > 0$ for $c_3 \neq 0$ we can conclude that there is no eigenfunction of the form (22).

If $\lambda > 0$, then the eigenfunction φ must satisfy (23) and (26). Then we obtain

$$0 = \tilde{\varphi}(1) - \tilde{\varphi}(0) = c_1(\cos(a) - 1) + c_2 \, \sin(a), \tag{27}$$

$$0 = \tilde{\varphi}'(1) - \tilde{\varphi}'(0) = a \, \left(c_2(\cos(a) - 1) - c_1 \, \sin(a)\right). \tag{28}$$

Both equations (27) and (28) are satisfied for $a = 2\,k\,\pi$ with $k \in \mathbb{Z}$. For $a \neq 2\,k\,\pi$, equation (27) implies

$$c_1 = -c_2 \, \frac{\sin(a)}{\cos(a) - 1}.$$

Plugging this in (28) yields

$$\begin{aligned} 0 &= c_2 \, \left(\cos(a) - 1 + \frac{\sin(a)^2}{\cos(a) - 1}\right) \\ &\iff 0 = 1 - 2\cos(a) + \cos(a)^2 + \sin(a)^2 = 2 - 2\cos(a), \end{aligned}$$

which implies $1 = \cos(a)$ and thus the contradiction $a = 2\,k\,\pi$. Hence the eigenfunctions can be only of the form

$$\varphi(s) = c_1 \, \cos(as) + c_2 \, \sin(as) + c_3$$

with $a = 2\,k\,\pi$ and $k \in \mathbb{Z}$. Since $a^2 := \frac{2}{|\lambda|} = \frac{2}{\lambda}$ we have $\lambda = \frac{2}{(2k)^2\,\pi^2} \neq \frac{1}{12}$ for all $k \in \mathbb{N}$. Hence either $c_1 = 0 = c_2$ or $c_3 = 0$ must be satisfied. In the case of $c_3 = 0$ we can restrict ourselves to $k \in \mathbb{N}$ because of symmetry and we get that the functions

$$\left\{\sqrt{2}\, \cos(2k\,\pi\, s);\; k \in \mathbb{N}\right\} \cup \left\{\sqrt{2}\, \sin(2k\,\pi\, s);\; k \in \mathbb{N}\right\}$$

are orthogonal and normalized. These functions are orthogonal to the constant function c_3 which should satisfy $c_3 = 1$ to be normalized. □

Proof of Theorem 2
a) Properties (9), (11) and Lemma 1 provide

$$\begin{aligned} & E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,Z_2=z_2)-\frac{1}{4} \\ = & \left(\tau(r_1,r_2)-\frac{1}{2}\right)\left(\frac{1}{2}-|G(t_1)-G(t_2)|\right) \\ = & -\frac{1}{2}\,\varphi_*(r_1)\,\varphi_*(r_2)\,\sum_{l=1}^{\infty}\lambda_l\,\varphi_l(G(t_1))\,\varphi_l(G(t_2)) \\ = & \sum_{l=1}^{\infty}-\frac{1}{2}\,\lambda_l\,\varphi_*(r_1)\,\varphi_l(G(t_1))\,\varphi_*(r_2)\,\varphi_l(G(t_2)). \end{aligned}$$

Since the residuals R_n and the explanatory variables T_n are independent, the functions $\widehat{\varphi}_l(y,t) := \varphi_*(y-x(t)^\top\beta)\ \varphi_l(G(t))$ with $l \in \mathbb{N}$ are orthogonal and normalized. Hence $\widehat{\lambda}_l := -\frac{1}{2}\,\lambda_l$ with $\widehat{\lambda}_{2l-1} = \frac{-1}{\pi^2\,(2l-1)^2}$ and $\widehat{\lambda}_{2l} = \frac{-1}{\pi^2\,(2l-1)^2}$ for $l \in \mathbb{N}$ are the eigenvalues of the spectral decomposition of $E(\psi_\beta(Z_1,\ldots,Z_{q+1})|Z_1=z_1,Z_2=z_2)-\frac{1}{4}$ so that the assertion follows from (7).
b) This assertion follows completely analogous to that in a) by using Lemma 2 instead of Lemma 1. □

Acknowledgement

This research was done while the author was visiting the Department of Mathematical and Statistical Sciences of the University of Alberta, Edmonton, Canada. She wants to thank the department and Ivan Mizera for the hospitality. In particular she thanks Ivan Mizera for his interest, fruitful discussions and helpful hints. She is also very grateful for the comments of the referees which improved the paper.

References

[1] Arcones, M.A., Chen, Z. and Giné, E. (1994). Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.* **22**, 1460-1477.

[2] Brauer, F. and Nohel, J.A. (1968). *Elementary Differential Equations: Principles, Problems, and Solutions.* W.A. Benjamin, New York.

[3] Christmann, A. and Rousseeuw, P.J. (2001). Measuring overlap in logistic regression. *Computational Statistics and Data Analysis* **28**, 65-75.

[4] Donoho, D.L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20**, 1803-1827.

[5] Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.* **14**, 119-128.

[6] Farebrother, R.W. (1984). The distribution of a linear combination of central χ^2 random variables. A remark on AS 153: Pan's procedure for the tail probabilities of the Durbin-Watson statistic. *Applied Statistics* **33**, 363-366.

[7] Hubert, M. and Rousseeuw, P.J. (1998). The catline for deep regression. *J. Multivariate Anal.*, **66**, 270-296.

[8] Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419-426.

[9] Kamke, E. (1947). *Differentialgleichungen reeller Funktionen.* Chelsea Publishing Company, New York.

[10] Lee, A.J. (1990). *U-Statistics. Theory and Practice.* Marcel Dekker, New York.

[11] Liu, R.Y. (1988). On a notion of simplicial depth. *Proc. Nat. Acad. Sci. USA* **85**, 1732-1734.

[12] Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18**, 405-414.

[13] Liu, R.Y. (1992). Data depth and multivariate rank tests. In *L_1-Statistical Analysis and Related Methods*, ed. Y. Dodge, North-Holland, Amsterdam, 279-294.

[14] Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88**, 252-260.

[15] Mizera, I. (2002). On depth and deep points: A calculus. *Ann. Statist.* **30**, 1681-1736.

[16] Mizera, I. and Müller, Ch.H. (2003). Location-scale depth. *To appear in J. Amer. Statist. Assoc.*

[17] Mosler, K. (2002). *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach.* Lecture Notes in Statistics **165**, Springer, New York.

[18] Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

[19] Rousseeuw, P.J. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing* **8**, 193-203.

[20] Rousseeuw, P.J. and Hubert, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94**, 388-433.

[21] Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians*, Vancouver 1974, **2**, 523- 531.

[22] Van Aelst, S. and Rousseeuw, P.J. (2000). Robustness of deepest regression. *J. Multivariate Anal.* **73**, 82-106.

[23] Van Aelst, S., Rousseeuw, P.J., Hubert, M. and Struyf, A. (2002). The deepest regression method. *J. Multivariate Anal.* **81**, 138-166.

[24] Witting, H. and Müller-Funk, U. (1995). *Mathematische Statistik II.* Teubner, Stuttgart.

[25] Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *Ann. Statist.* **28**, 461-482.

[26] Zuo, Y. and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28**, 483-499.

mueller@math.uni-oldenburg.de
Institute of Mathematics
Carl von Ossietzky University Oldenburg
Postfach 2503
D-26111 Oldenburg
Germany