Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS

Radka Sabolová

Statistical inference based on saddlepoint approximations

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis:	prof. RNDr. Jana Jurečková DrSc.
Study programme:	Mathematics
Specialization:	Probability and Mathematical Statistics

Prague 2014

I would like to thank to my supervisor prof. Jana Jurečková for her help, valuable advice and support during my doctoral studies. I am also very thankful to prof. Elvezio Ronchetti (University of Geneva, Switzerland) and assoc. prof. Milan Stehlík (Johannes Kepler University, Linz, Austria).

The work was supported by the grant SVV-2014-260105.

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague, 30.05.2014

Název práce: Statistická inference založená na aproximací pomocí metody sedlového bodu

Autor: Radka Sabolová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí disertační práce: prof. RNDr. Jana Jurečková, DrSc., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Metody založené na aproximacích pomocí sedlového bodu pro M-odhady se ve statistice osvědčily jako velmi přesné a robustní i pro výběry malého rozsahu. V této práci byly touto metodou odvozeny aproximace hustoty a testy v kvantilové regresi, a to pro parametrický i neparametrický případ. Kromě toho byl navržen i test o hodnotě regresního kvantilu založený na asymptotickém rozdělení zprůměrovaných regresních kvantilů. Tyto testy byly pak porovnány s jinými dostupnými testy v simulační studii. Poslední část práce je věnována speciálnímu případu Kullback-Leiblerovy divergence pro exponenciální rodinu rozdělení, na základě které byly také odvozeny aproximace hustoty maximálně věrohodného odhadu a suficientní statistiky pomocí metody sedlového bodu.

Klíčová slova: metoda sedlového bodu, kvantilová regrese, I-divergence

Title: Statistical inference based on saddlepoint approximations

Author: Radka Sabolová

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Jana Jurečková, DrSc., Department of Probability and Mathematical Statistics

Abstract: The saddlepoint techniques for M-estimators have proved to be very accurate and robust even for small sample sizes. Based on these results, saddlepoint approximations of density of regression quantile and saddlepoint tests on the value of regression quantile were derived, both in parametric and nonparametric setup. Among these, a test on the value of regression quantile based on the asymptotic distribution of averaged regression quantiles was also proposed and all these tests were compared in a numerical study to the classical tests. Finally, special case of Kullback-Leibler divergence in exponential family was studied and saddlepoint approximations of the density of maximum likelihood estimator and sufficient statistic were also derived using this divergence.

Keywords: saddlepoint method, quantile regression, I-divergence

Contents

Introduction

Once one derives an estimator or a test statistic, it is also important to derive its distribution to construct a test or confidence intervals. One may rely on asymptotics and with the aid of central limit theorem approximate the distribution.

However, this approach has certain drawbacks. This approximation (under some assumptions) yields normal distribution, even for the estimators that are heavy-tailed under finite number of observations. The asymptotics work very well for larger sample sizes but it is desirable to have asymptotics that work even for $n = 1$ (see Huzurbazar (1999)). In case of a smaller number of observations, one may consider using Edgeworth expansions in order to improve the approximation to the normal distribution. The ordinary central limit theorem leads only to the first term of the expansion. One or two more terms often improve this basic approximation and provide more insight into the distribution. Moreover, these results help us to study the procedures equivalent to first order and discriminate among them (see Bickel (1974)).

Unfortunately, the approximation by Edgeworth expansion might yield negative values for the density in the tails. One of the methods overcoming this drawback is the so called saddlepoint method with the relative error $O(n^{-1})$. The density of the estimator is rewritten using Fourier transform as an integral of a function of the cumulant generating function. Then, a saddlepoint of this cumulant generating function is found and the integration path is modified in a way that it goes through the path of steepest descent from the saddlepoint, so it captures most of the mass and the remaining contributions to the integral become negligible (for more information see Huber and Ronchetti (2009)).

The saddlepoint approximations were introduced into statistics by Daniels (1954) where approximation to the density of mean was derived. Since then these techniques have been further elaborated by Hampel (1973) and by Field and Hampel (1982). For the overview on saddlepoint techniques see Field and Ronchetti (1990), Jensen (1995), and Butler (2007). Hampel (1973) proposed the *small sample asymptotics* where he suggested expanding the function $\frac{g_n'}{g_n}$ instead of expanding g_n to gain more precise approximation even for very small sample sizes.

Based on the success and accuracy of saddlepoint approximation to the density for the estimators (see Jurečková and Sabolová (2011)), saddlepoint test for M-estimator in parametric model was proposed in Ronchetti et al. (2003). These tests are asymptotically χ^2-distributed but under finite number of observations exhibit better properties than classical tests based on likelihood (Wald, Rao score, likelihood ratio). In Ronchetti et al. (2003) these tests were also extended for a nonparametric setting where the distribution of the observations under null hypothesis is not available. These procedures exhibit better finite sample behavior than the available tests by combining excellent accuracy, even in small samples, and robustness. This thesis will mostly deal with saddlepoint approximations and tests in quantile regression. Regression quantiles are M-estimators with not very complicated ρ-function that yields explicit formulae for saddlepoint approximations and test statistics, that is usually not the case for saddlepoint methods.

The thesis is organized as follows. The first chapter contains overview of basic results in the field of saddlepoint approximations and tests. The saddlepoint approximations for the density are compared to the exact density for a special class of estimators in a numerical study. In the second chapter, the density of simple quantiles and regression quantiles is derived using saddlepoint approximations in two cases, whether the density of the errors in the linear model is specified or unspecified. The procedure of approximating the density of averaged regression quantiles is outlined. The third chapter deals with tests on the value of simple and regression quantile. A test based on averaged regression quantiles is introduced. The saddlepoint tests for regression quantiles are derived (both in parametric and in nonparametric case) and the validity of the method is demonstrated through a simulation study, which shows both the robustness and the great accuracy of the new test compared to the available alternatives. In the last chapter the Kullback-Leibler divergence for exponential family of distributions is considered and saddlepoint approximations based on this approach are derived.

Chapter 1

Saddlepoint techniques in statistics

In this chapter theory of saddlepoint approximations for the density of multivariate M-estimators will be presented. Performance of these approximations will be studied numerically for equivariant M-estimators, i.e. a special class of estimators for which a formula for density under the finite number of observations is available. Theory of saddlepoint approximations for the density of functions of M-estimators will be also briefly outlined, and in the following chapter applied in order to approximate the density of regression quantiles and averaged regression quantiles. Later, the general theory of saddlepoint tests for M-estimators will be also introduced in parametric and also nonparametric setup.

1.1 Saddlepoint approximations for the density of estimators

Let $X_1, \ldots, X_n$ be i.i.d. vectors with the distribution $F(x, \theta)$ and the corresponding density $f(x, \theta)$. Consider an M-estimator T_n of parameter θ, defined as a solution of the equation

$$\sum_{i=1}^{n} \psi(X_i, t) = 0$$

for a suitable ψ-function.

Daniels (1954) derived the saddlepoint approximation of the density of mean. The density was rewritten using Fourier transform and the integration path was modified to go through the path of steepest descent, so it captured most of the mass around saddlepoint.

An approximation of density of an M-estimator is based on techniques derived for the mean (Field and Ronchetti (1990), Chapter 3); T_n is expressed as a mean up to a certain order and then the saddlepoint approximation for the mean is used.

We will introduce a saddlepoint approximation for the density of a multivariate M-estimator. Denote D_j differentiation with respect to θ_j. The following assumptions on the functions ψ and $f(x, \theta)$ are required in order to develop the approximation. The smoothness conditions on ψ-function are necessary in order to make Taylor expansion. The saddlepoint approximation for the density of T_n will be denoted by $f_n(t)$. Further, we make the following assumptions:

(S1) The equation $\sum_{i=1}^{n} \psi_j(X_i, t) = 0$, for $j = 1, \dots, p$ has a unique solution T_n.

(S2) There is an open subset $U \subset \mathbb{R}^p$ such that
(i) $F_\theta(U) = 1$ for each $\theta \in \Theta$
(ii) the derivatives $D_j\ \psi_r(x, \theta)$, $D_k D_j \psi_r(x, \theta)$, $D_l D_k D_j \psi_r(x, \theta)$ exist for $1 \leq r, j, k, l \leq p$.

(S3) For each compact $K \subset \Theta$
(i) for $0 \leq j, k \leq p$,

$$\sup_{\theta_0 \in K} \mathsf{E}_{\theta_0} |D_k D_j \psi_r(X, \theta_0)|^4 < \infty,$$

(ii) there is an $\varepsilon > 0$ such that for $1 \leq r, j, k, l \leq p$

$$\sup_{\theta_0 \in K} \mathsf{E}_{\theta_0} (\max_{|\theta - \theta_0| \leq \varepsilon} |D_l D_k D_j \psi_r(X, \theta)|^3) < \infty.$$

(S4) for each $\theta_0 \in \Theta$

$$\mathsf{E}_{\theta_0} \psi_r(X, \theta_0) = 0$$

and the matrices

$$\begin{aligned} A(\theta_0) =& \mathsf{E}_{\theta_0} \frac{\partial \psi}{\partial \theta}(X, \theta_0), \\ C(\theta_0) =& \mathsf{E}_{\theta_0} [\psi(X, \theta_0) \psi_r^T(X, \theta_0)] \end{aligned}$$

are non singular.

(S5) The functions $A(\theta)$ and $\mathsf{E}_\theta[(D_{k_1} D_{j_1} \psi_{r_1})(D_{k_2} D_{j_2} \psi_{r_2})]$, $0 \leq j_1, j_2, k_1, k_2 \leq p$, $k_1 + j_1 \geq 1$, $k_2 + j_2 \geq 1$, $1 \leq r_1, r_2 \leq p$ are continuous on Θ.

The following theorem summarizes the approximation $f_n(t)$ (for proof, see Field and Ronchetti (1990)):

Theorem 1. *If T_n represents the solution of $\sum_{i=1}^{n} \psi_r(x_i, t) = 0$, $r = 1, \dots, p$, and Assumptions (S1) – (S5) are satisfied, then an asymptotic expansion for the density of T_n, say f_n, is*

$$f_n(t_0) = (n/2\pi)^{p/2} c^{-n}(t_0) |detA| |det\Sigma|^{-1/2} \{1 + O(1/n)\},$$

where p-dimensional vector $\alpha(t_0)$ is a solution of

$$\int \psi_r(x, t_0) \exp\left\{\sum_{j=1}^{p} \alpha_j \psi_j(x, t_0)\right\} f(x) dx = 0, \qquad r = 1, \dots, p, \tag{1.1}$$

$$c^{-1}(t_0) = \int \exp\left\{\sum_{j=1}^{p} \alpha_j(t_0)\psi_j(x,t_0)\right\} dx$$

$$A = \left\{E\frac{\partial\psi(x,t)}{\partial t_r}|_{t=t_0}\right\}_{1\leq r,j\leq p}$$

$$\Sigma = \{E\psi_j(x,t_0)\psi_r(x,t_0)\}_{1\leq r,j\leq p}$$

and all expectations are with respect to the conjugate density

$$h_{t_0}(x) = c(t_0)\exp\left\{\sum_{j=1}^{p} \alpha_j\psi_j(x,t_0)\right\} f(x).$$

The error term holds uniformly for all t_0 in a compact set.

Theorem 1 gives the following approximation of the density f_n :

$$g_n(t_0) = \sqrt{\frac{n}{2\pi}} c^{-n}(t_0)\frac{A(t_0)}{\sigma(t_0)}. \tag{1.2}$$

Notice that the saddlepoint approximation requires the existence of moment generating function of ψ-function. For bounded functions ψ this condition is satisfied even for random variables X having heavy-tailed distribution. In the case of the approximation to the density of mean, the existence of moment generating function of a random variable X is required.

Remark 1. *For normally distributed random variables, the* (1.1) *can be solved analytically and $\alpha(t) = t$. For other distributions iterative methods (e.g., Newton-Raphston) have to be used in order to find the solution of equation* (1.1). *When looking for saddlepoint, one might use the fact that $\alpha(0) = 0$ as a starting point.*

We are often interested not in the approximation of the density of the estimator, but we are looking for the approximation of the function of estimator. The saddlepoint approximation of the function of an M-estimator was introduced in Fan and Field (1995) where a saddlepoint approximation to the marginal density of a general function $u(.)$ of an M-estimator was derived. They suggested the following procedure:

- center the joint conjugate density to the point r_0 in the expectation, where $r_0 = u(\theta_0)$
- $\hat{u} = u(\hat{\theta})$ is approximated by a linear combination of score functions
- the approximation is transformed back to an approximation under original density f.

1.2 Numerical comparison to the exact density of equivariant M-estimators

In some special situations, even exact density of the estimator of the shift parameter can be derived. One of these cases is a translation equivariant M-estimator, i. e. T_n satisfying

$$T_n(X_1 + c, \ldots, X_n + c) = T_n(X_1, \ldots, X_n) + c.$$

1.2.1 Density of equivariant M-estimators

Let $X_1, \ldots, X_n$ be a sample from the distribution with distribution function $F(x - \theta)$ such that F has an absolutely continuous density f and finite Fisher information. Let $S_n = S_n(X_1, \ldots, X_n)$ be a statistic whose distribution function $H_\theta(s)$ is continuously differentiable in θ. Then we have the following identity for the derivative of $H_\theta(s)$ w.r.t. θ (see Jurečková (1999)):

$$\begin{aligned}\frac{\partial H_\theta(s)}{\partial \theta} &= \int_{S(x_1,\ldots,x_n)\le s} \cdots \int \sum_{i=1}^{n} \left(-\frac{f'(x_i-\theta)}{f(x_i-\theta)}\right) \prod_{i=1}^{n} f(x_i-\theta)dx_1\ldots dx_n \\ &= \mathsf{E}_\theta \left[\sum_{i=1}^{n} \left(-\frac{f'(x_i-\theta)}{f(x_i-\theta)}\right) .I[S(X_1,\ldots,X_n)\le s]\right]. \qquad (1.3)\end{aligned}$$

Let especially, T_n be a translation equivariant estimator of θ and let $g_\theta(t)$ be its density. A possible finite-sample expression for the density $g_\theta(t)$ we get from (1.3); it will further enable to study various properties of T_n :

$$g_\theta(t) = \int_{T(x_1,\ldots,x_n)\le t} \cdots \int \sum_{i=1}^{n} \frac{f'(x_i-\theta)}{f(x_i-\theta)} \prod_{k=1}^{n} f(x_k-\theta)dx_1\ldots dx_n \quad (1.4)$$

$$= \mathsf{E}_0 \left\{ \sum_{i=1}^{n} \frac{f'(X_i)}{f(X_i)}\ I\Big[T(X_1,\ldots,X_n)\le t-\theta\Big]\right\}.$$

If T_n is a solution of the equation $\sum_{i=1}^{n} \psi(X_i - t) = 0$ with monotone ψ, then $g(t)$ can be rewritten

$$g_\theta(t) = \mathsf{E}_0 \left\{ \sum_{i=1}^{n} \frac{f'(X_i)}{f(X_i)}\ I\Big[\sum_{j=1}^{n} \psi(X_j - (t-\theta)) \le 0\Big]\right\}. \qquad (1.5)$$

Although (1.5) is an explicit formula for the density, the integral has to be approximated numerically, therefore the applicability of the formula is limited for larger sample sizes. In (Jurečková and Sabolová, 2011) we studied the performance of saddlepoint approximation to the density and compared it with the exact density for the following translation equivariant M-estimators:

(i) the mean (M-estimator with $\psi(x) = x$),

(ii) the Huber M-estimator with

$$\psi(x) = \begin{cases} x & : |x| \le k \\ k \cdot \mathrm{sign}(x) & : |x| \ge k \end{cases},$$

where k was set to 1.4,

(iii) the maximum likelihood estimator, i.e. the M-estimator with

$$\psi(x) = -\frac{f'(x)}{f(x)}.$$

The calculation was made for the following parent distributions:

- Standard normal distribution N(0, 1) with density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $-\frac{f'(x)}{f(x)} = x$.
- Logistic distribution Log(0, 1) with density $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$ and $-\frac{f'(x)}{f(x)} = -\frac{1-e^x}{1+e^x}$.
- The Student t-distribution with 3 degrees of freedom t_3 with density $f(x) = \frac{6\sqrt{3}}{\pi(3+x^2)^2}$ and $-\frac{f'(x)}{f(x)} = \frac{4x}{3+x^2}$.
- The Student t-distribution with 5 degrees of freedom t_5 with density $f(x) = \frac{100\sqrt{5}}{3\pi(5+x^2)^3}$ and $-\frac{f'(x)}{f(x)} = \frac{6x}{5+x^2}$.

Various steps and regions of numerical integration depending on the shape of the distribution were considered for the specific densities. All numerical integrations were coded in `C`. As the approximation of multidimensional integrals is quite time consuming, the densities using formula (1.4) were evaluated for small n, namely for $n \le 4$. When approximating mean for $n = 1$ we get approximation of the original density f. Results of these approximations are presented in the Fig. 1.1, where exact density is drawn by the black solid line, saddlepoint approximation by the red dashed line and formula (1.4) by the green dotted line.

When approximating the density of T_n by formula (1.2), it is necessary to find $\alpha(t)$, which solves the equation

$$\int \psi(x,t) e^{\alpha(t)\psi(x,t)} f(x,\theta) dx = 0.$$

For $\psi(x,t) = x - t$ (i.e. the mean), and $f(x)$ being density of $N(0,1)$ is $\alpha(t)$ is equal to t. For other distributions and estimators one can use the fact that $\alpha(t) = 0$ for $t = 0$, which provides a good starting point for the Newton-Raphston method for other t. Then c, S and A are approximated by numerical integration and then inserted into formula (1.2) for f_n.

The results of the numerical study are plotted in Figures 1.2–1.12. Note that both methods lead to the similar outcomes for normally distributed data. The results are also quite similar for f logistic unless t is close to 0. The method based on (1.4) does not work for the Laplace distribution whose density is not continuously differentiable. As it was already observed by Field and Ronchetti (1990), their approximation does not work very well for the Laplace distribution, either. It was not possible to conduct a study for Cauchy distribution as the moment generating function does not exist. The differences between two methods are smaller for Huber estimator than for the mean of the Student's t-distribution, although the results improve. It is of interest that the approximations are very close to each other for the maximum likelihood estimators; this apparently demonstrates the important role of the score functions.

Even though formula (1.4) provides exact expression of the density of the estimator, the discrepancies are caused by approximation of integrals (even resulting in negative values for Laplace distribution). The main disadvantage of this method is its time complexity. Since for bigger n the computations take a lot of time, this method proves to be inefficient for big n. On the other hand, time consumed when approximating density by saddlepoint techniques does not grow for bigger n. As the simulations were done for very small sample sizes, the saddlepoint approximations proved to be very precise even in this extreme situation.

1.3 Saddlepoint test for M-estimators

As the saddlepoint approximations proved to be very precise, tests for M-estimators based on similar techniques were later also developed. The test statistic is based on the value of cumulant generating function of ψ-function evaluated at its saddlepoint. The test is first-order equivalent to classical tests based on likelihood, but exhibits better second-order properties (see Ronchetti et al. (2003)).

Let us consider a composite hypothesis

$$H_0 : \theta_1 = \theta_{1_0} \in \mathbb{R}^{p_1}, \theta_2 \in R^{p_2},$$

where $\theta = (\theta_1^T, \theta_2^T), \hat{\theta} = (\hat{\theta}_1^T, \hat{\theta}_2^T)$. We will shortly introduce general theory of saddlepoint tests for M-estimators, as developed in Ronchetti et al. (2003).

One dimensional test statistic

$$h(\hat{\theta}_1) = \inf_{\theta_2} \sup_{\lambda} \{-K_F(\lambda; (\hat{\theta}_1, \theta_2))\} \tag{1.6}$$

was proposed in Ronchetti et al. (2003), where

$$K_F(\lambda, \theta) = \log \mathsf{E}_F[e^{\lambda^T \psi(X_i, \theta)}]$$

is the cumulant generating function of the score $\psi(X_i, \theta)$ and expectation is computed with respect to the distribution of the observations under the null hypothesis. Note that (1.6) can be rewritten as

$$\sup_{\lambda} \{-K_F(\lambda; t)\} = -K_F(\lambda(t); t), \tag{1.7}$$

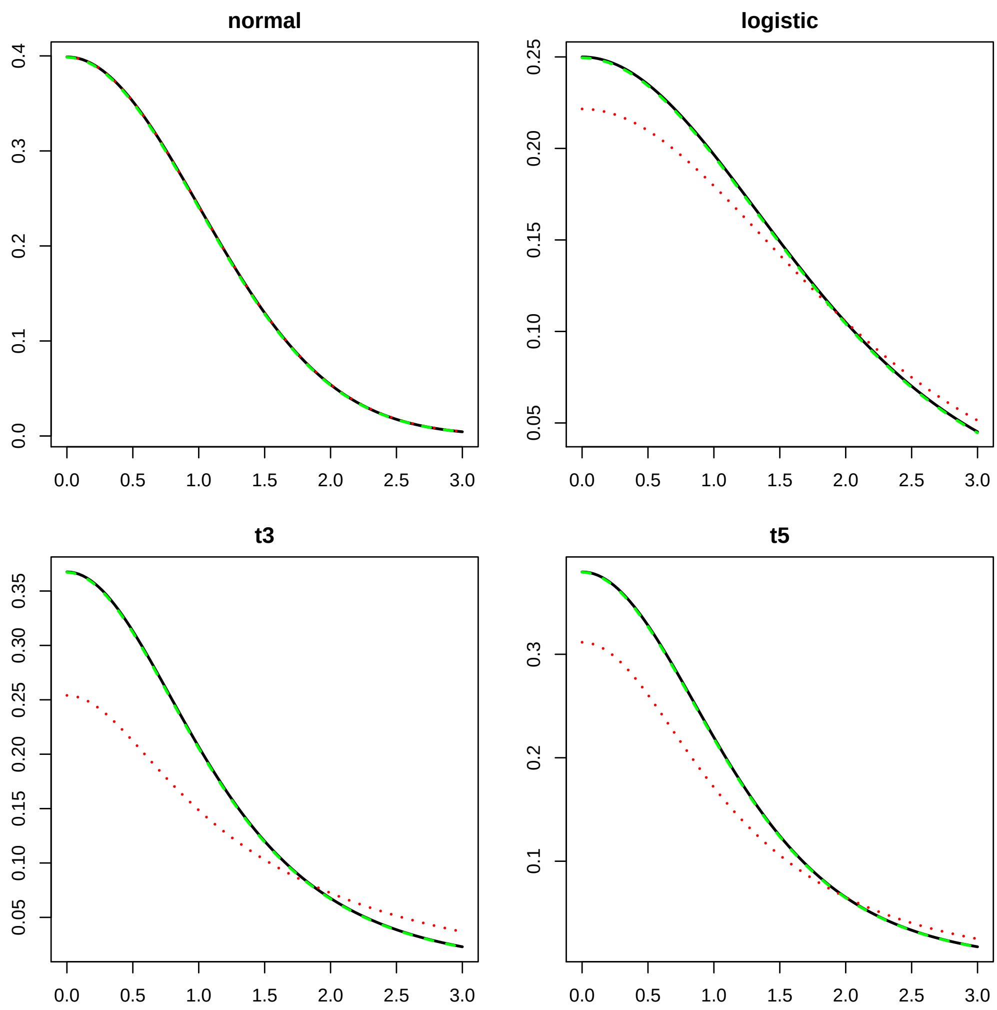


Figure 1.1: Estimated density compared to the exact one for different distributions, sample size $n = 1$

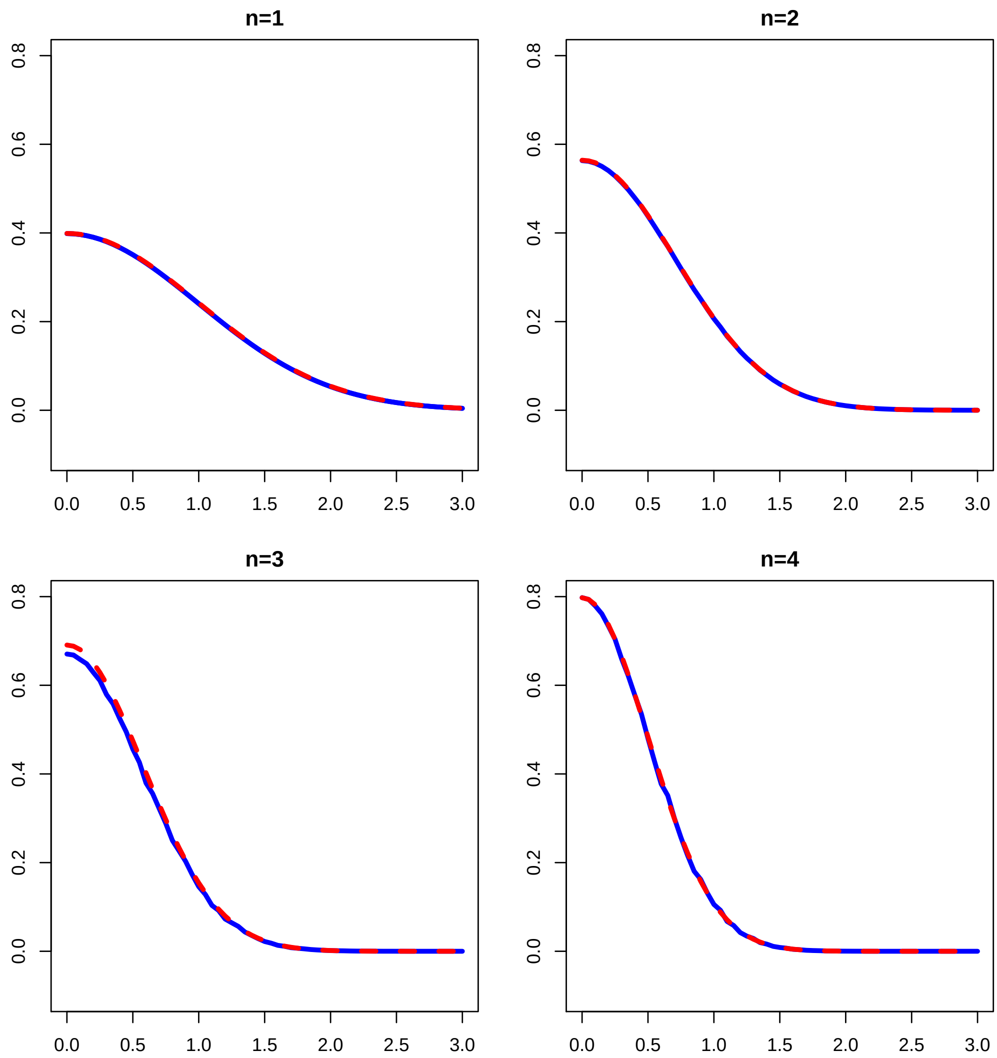


Figure 1.2: Comparison of exact density (blue line) of mean and its saddlepoint approximation (red dashed line), normal distribution

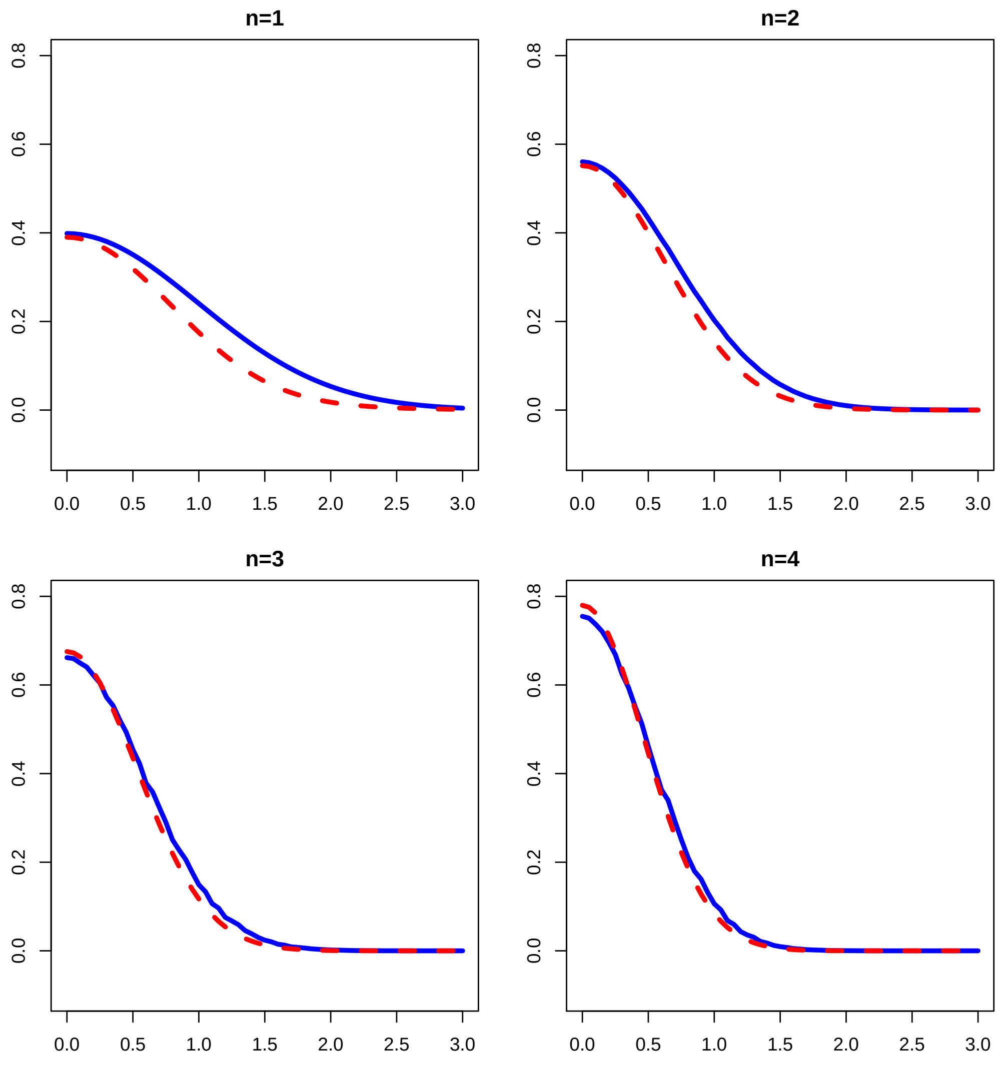


Figure 1.3: Comparison of exact density (blue line) of Huber estimator and its saddlepoint approximation (red dashed line), normal distribution

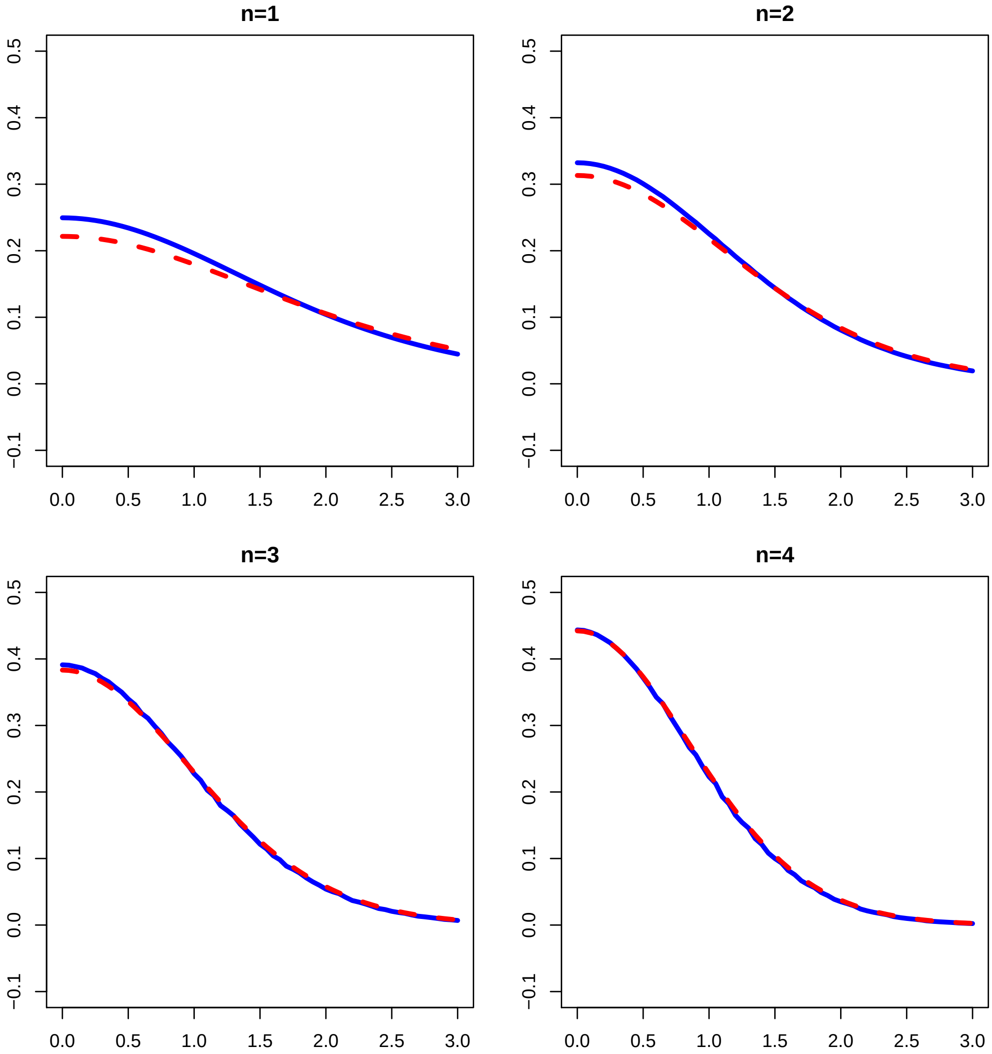


Figure 1.4: Comparison of exact density (blue line) of mean and its saddlepoint approximation (red dashed line), logistic distribution

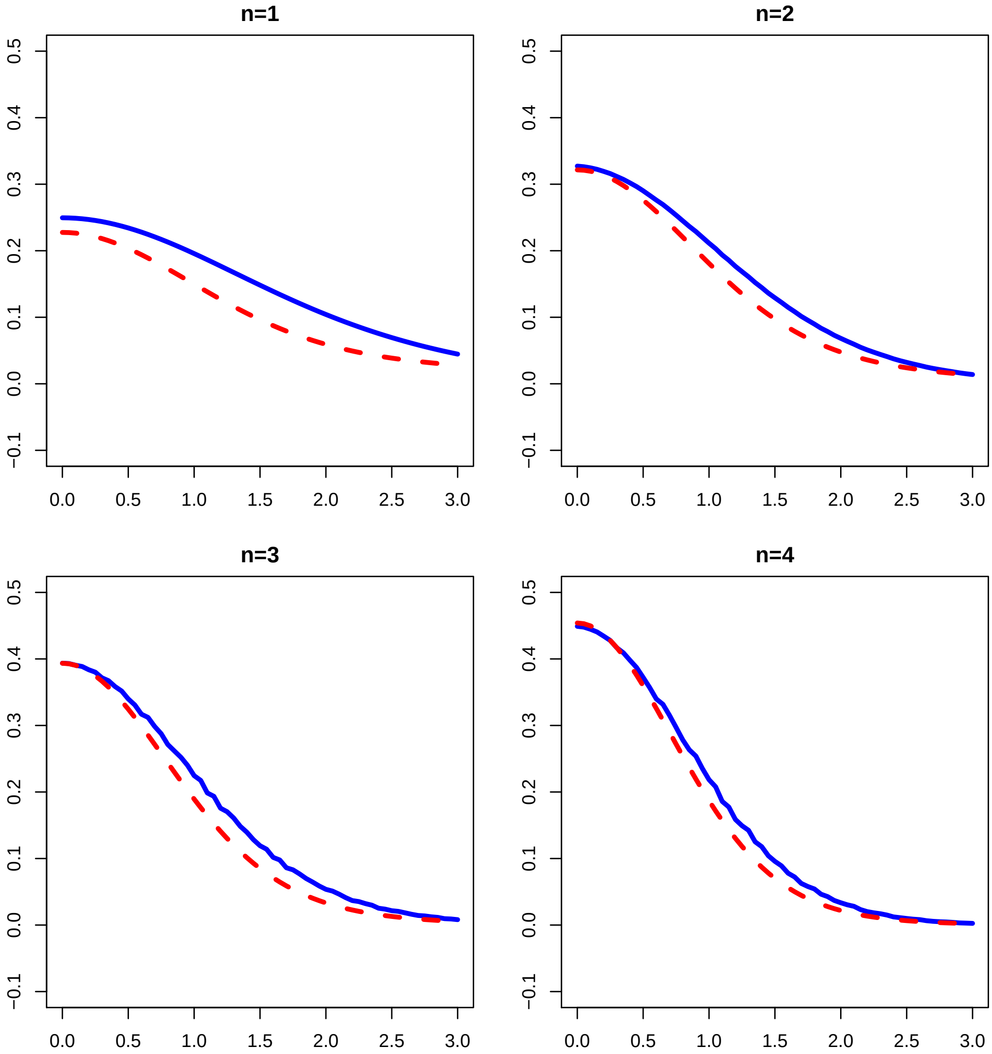


Figure 1.5: Comparison of exact density (blue line) of Huber estimator and its saddlepoint approximation (red dashed line), logistic distribution

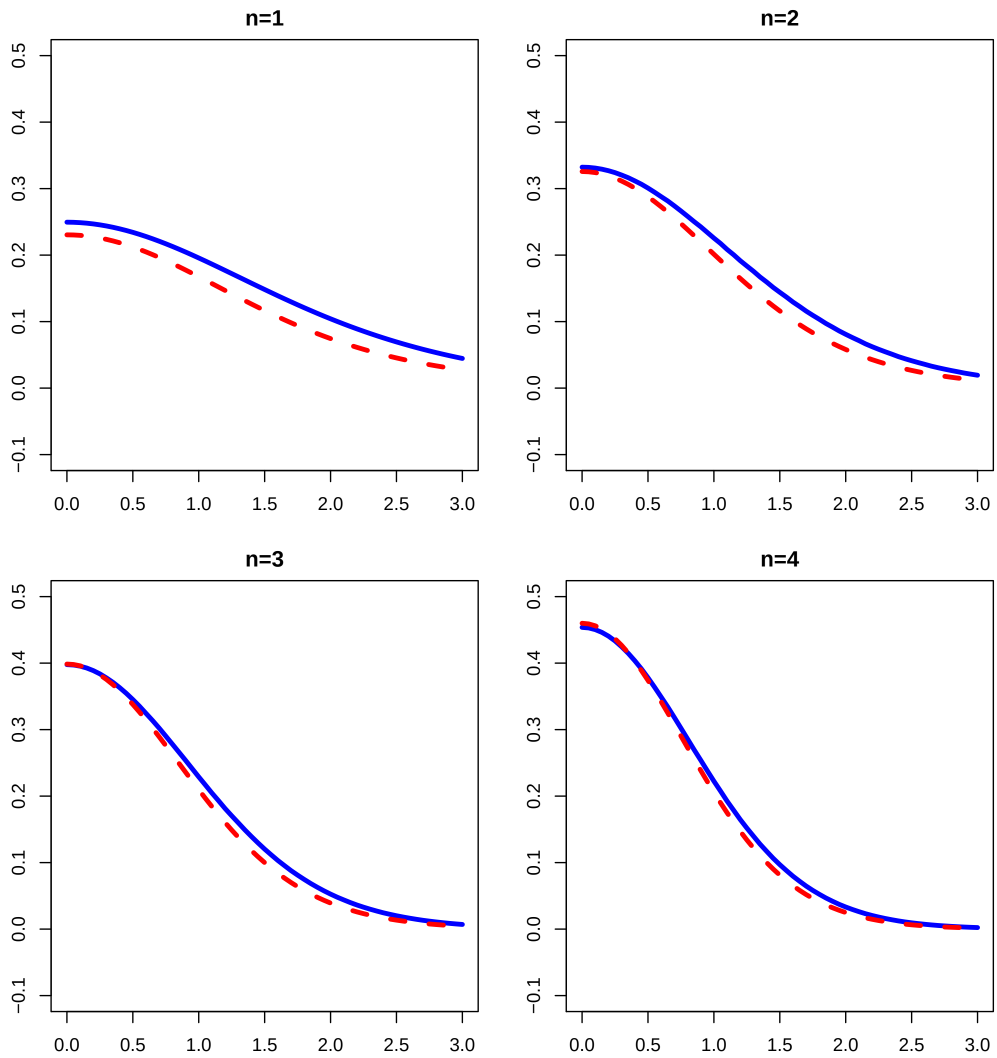


Figure 1.6: Comparison of exact density (blue line) of MLE and its saddlepoint approximation (red dashed line), logistic distribution

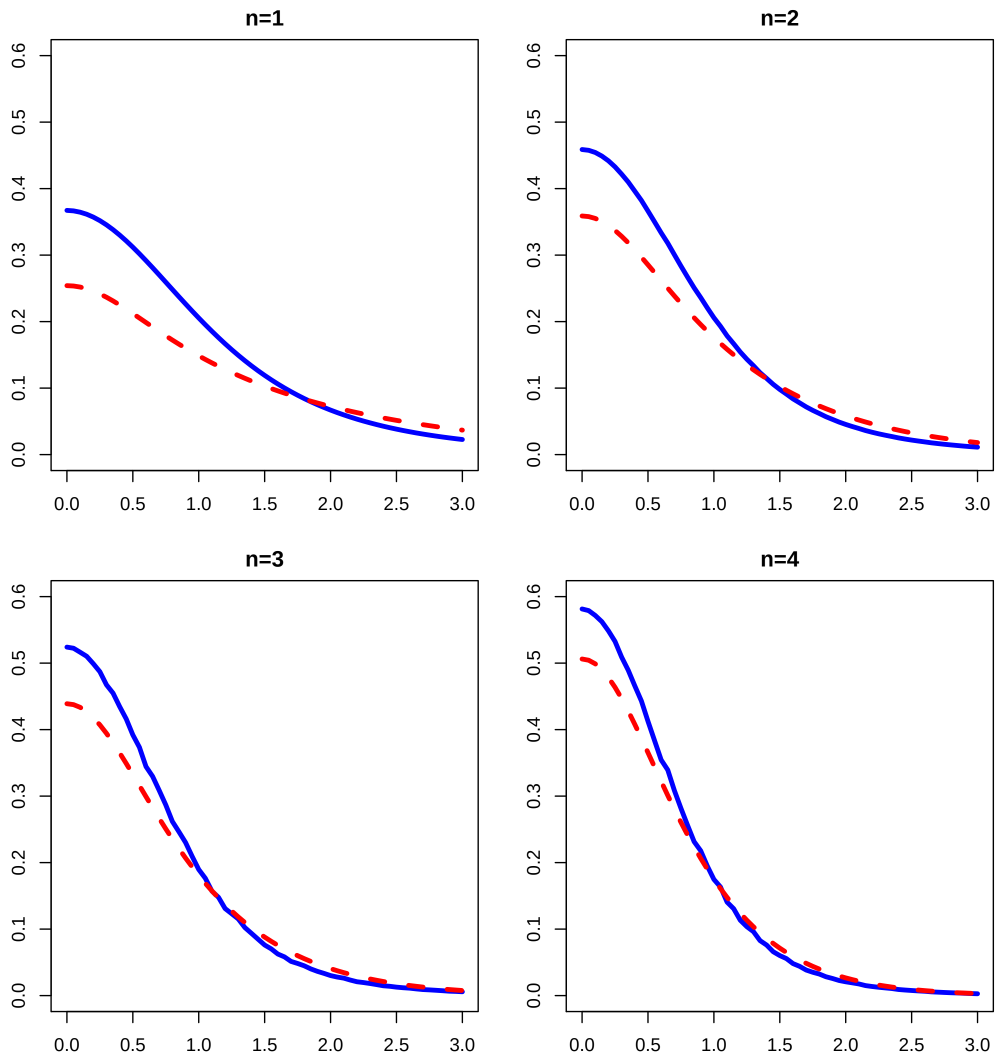


Figure 1.7: Comparison of exact density (blue line) of mean and its saddlepoint approximation (red dashed line), Student t_3 distribution

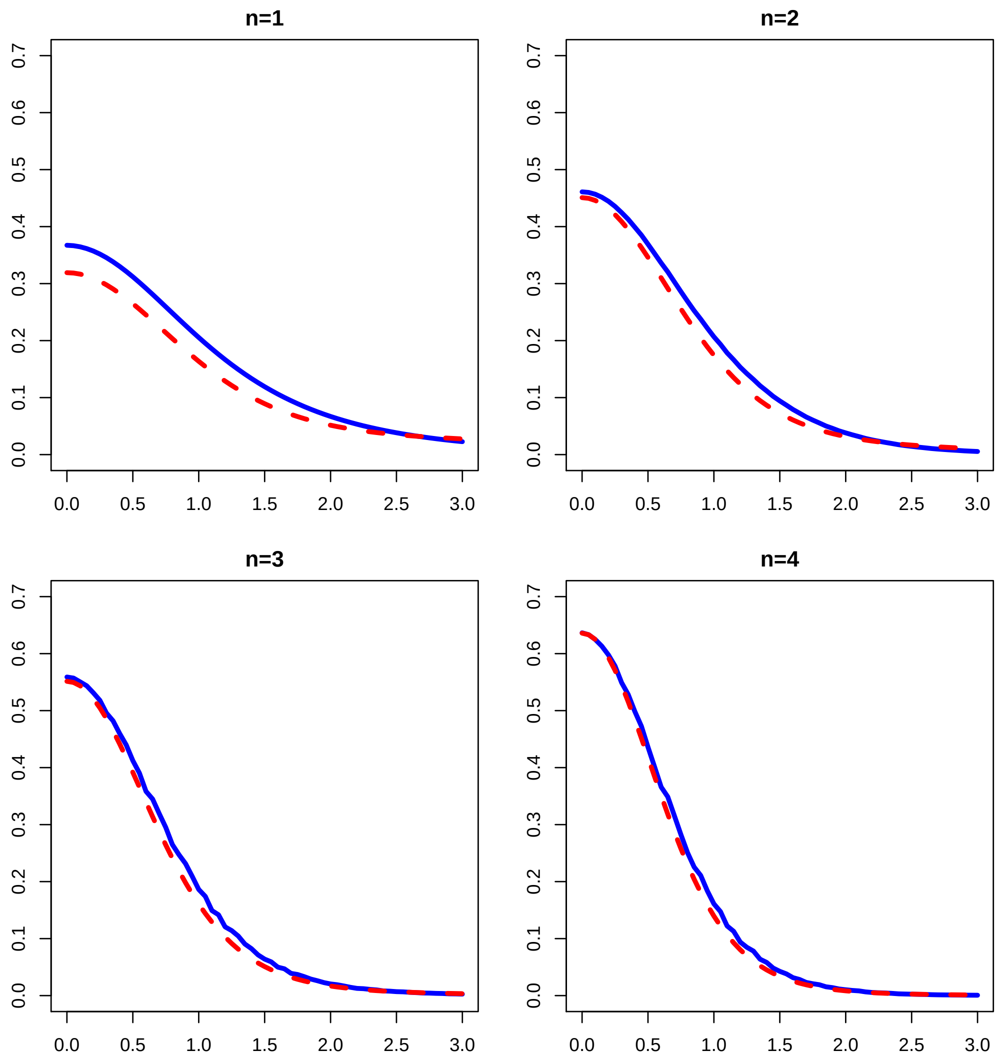


Figure 1.8: Comparison of exact density (blue line) of Huber estimator and its saddlepoint approximation (red dashed line), Student t_3 distribution

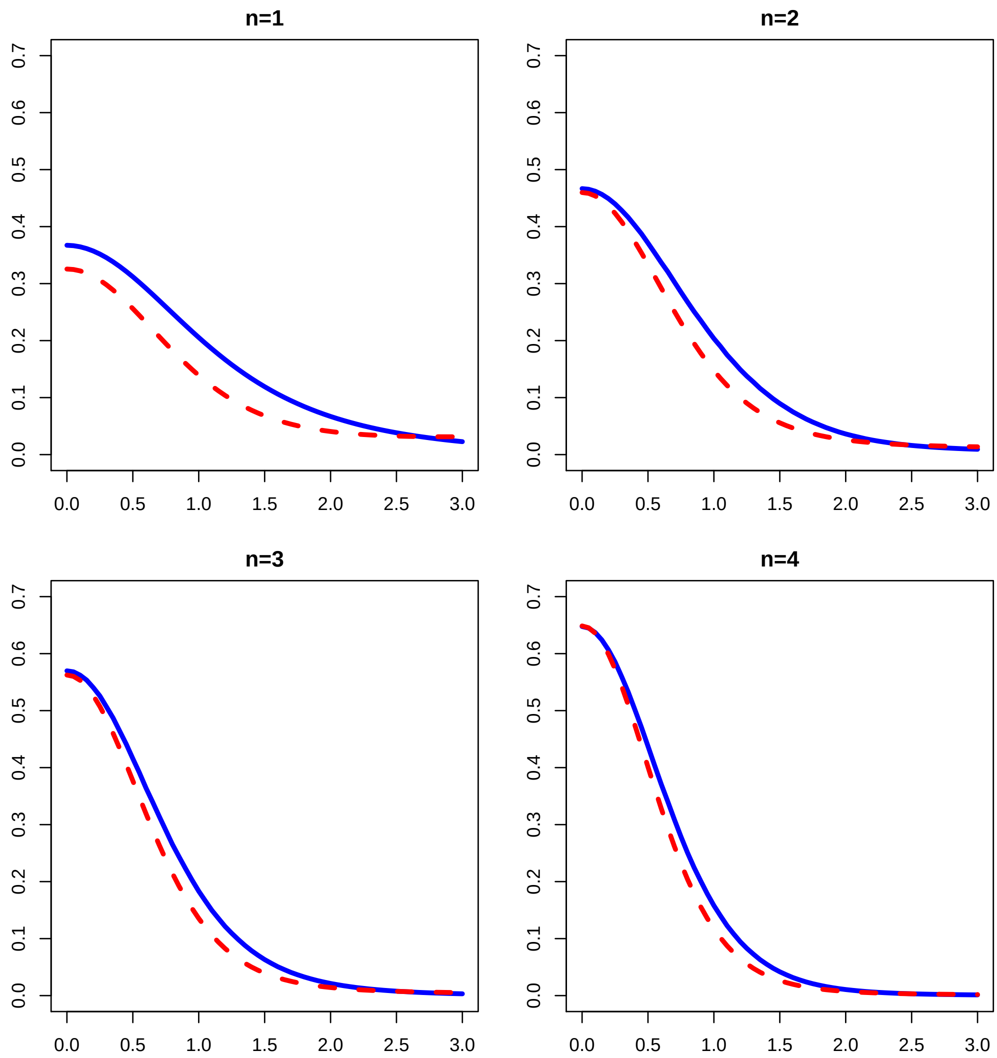


Figure 1.9: Comparison of exact density (blue line) of MLE and its saddlepoint approximation (red dashed line), Student t_3 distribution

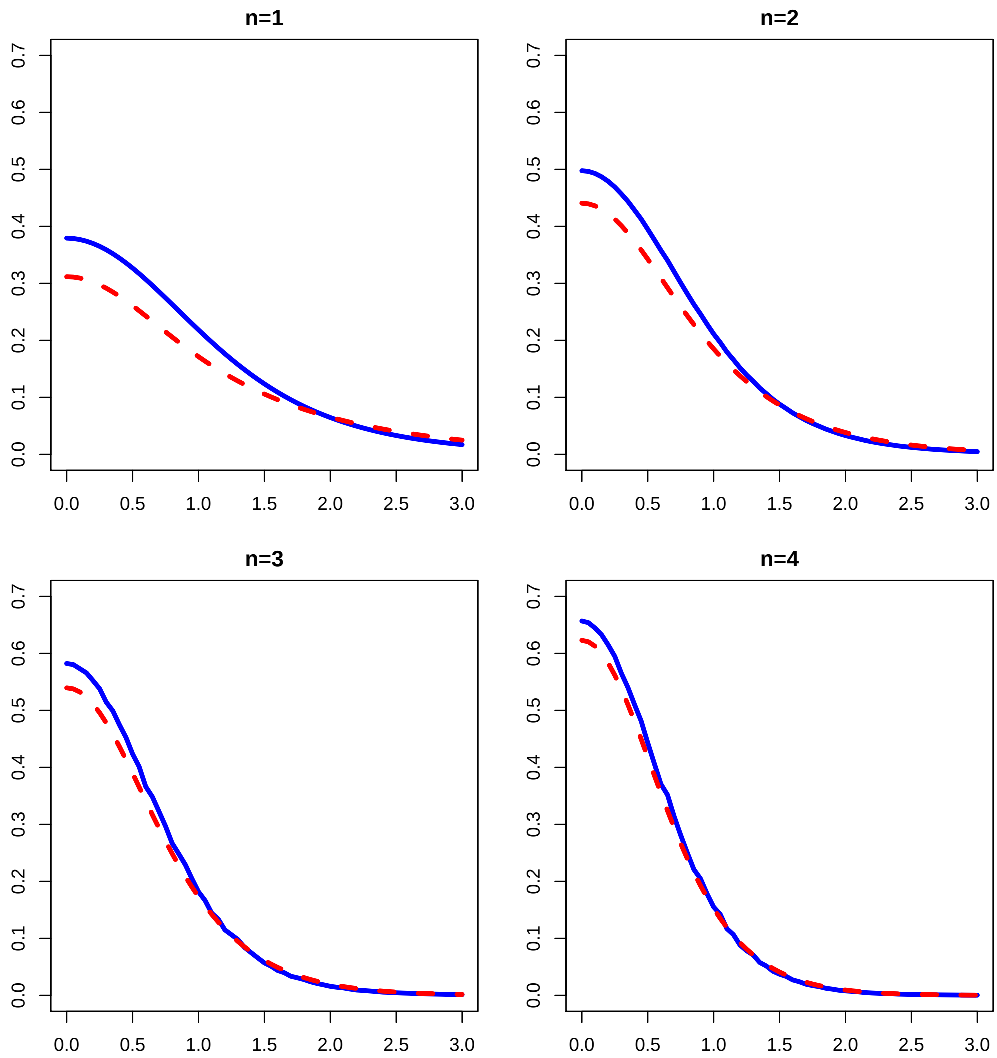


Figure 1.10: Comparison of exact density (blue line) of mean and its saddlepoint approximation (red dashed line), Student t_5 distribution

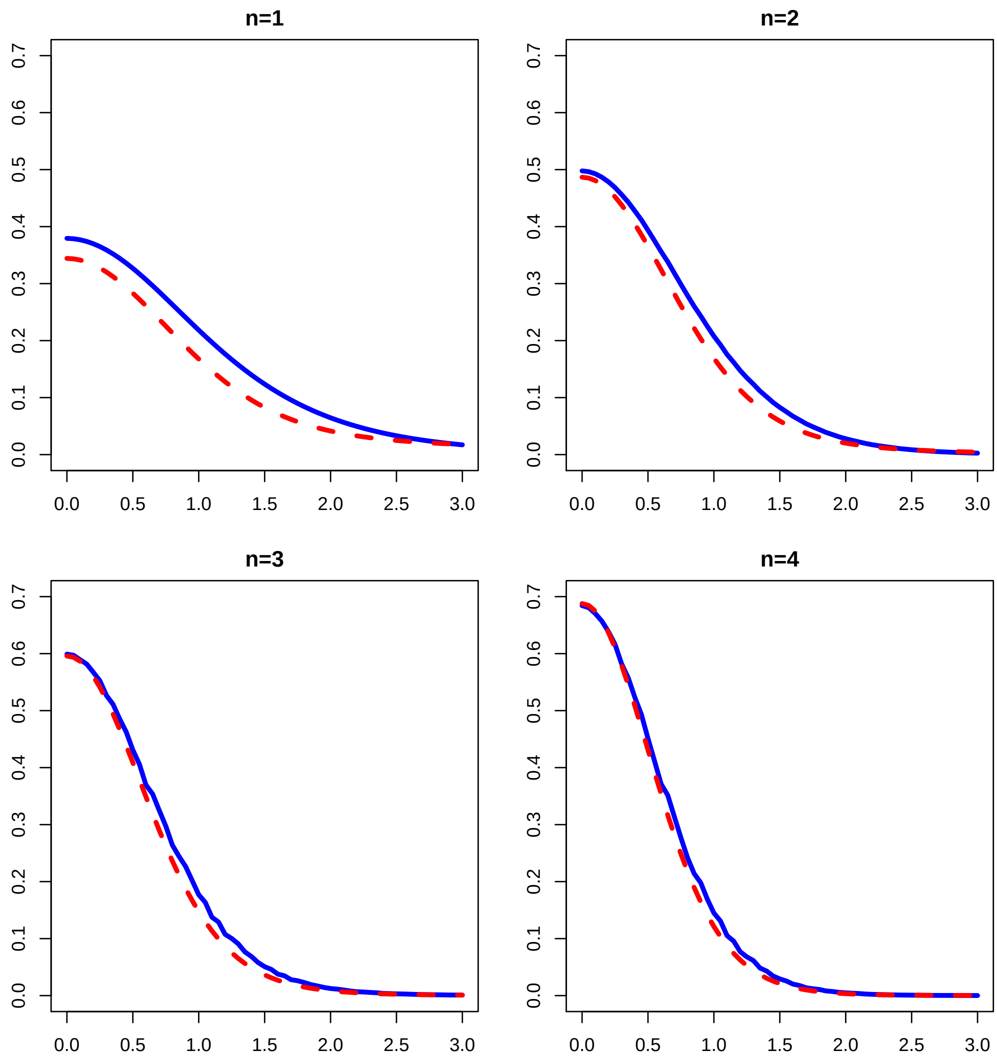


Figure 1.11: Comparison of exact density (blue line) of Huber estimator and its saddlepoint approximation (red dashed line), Student t_5 distribution

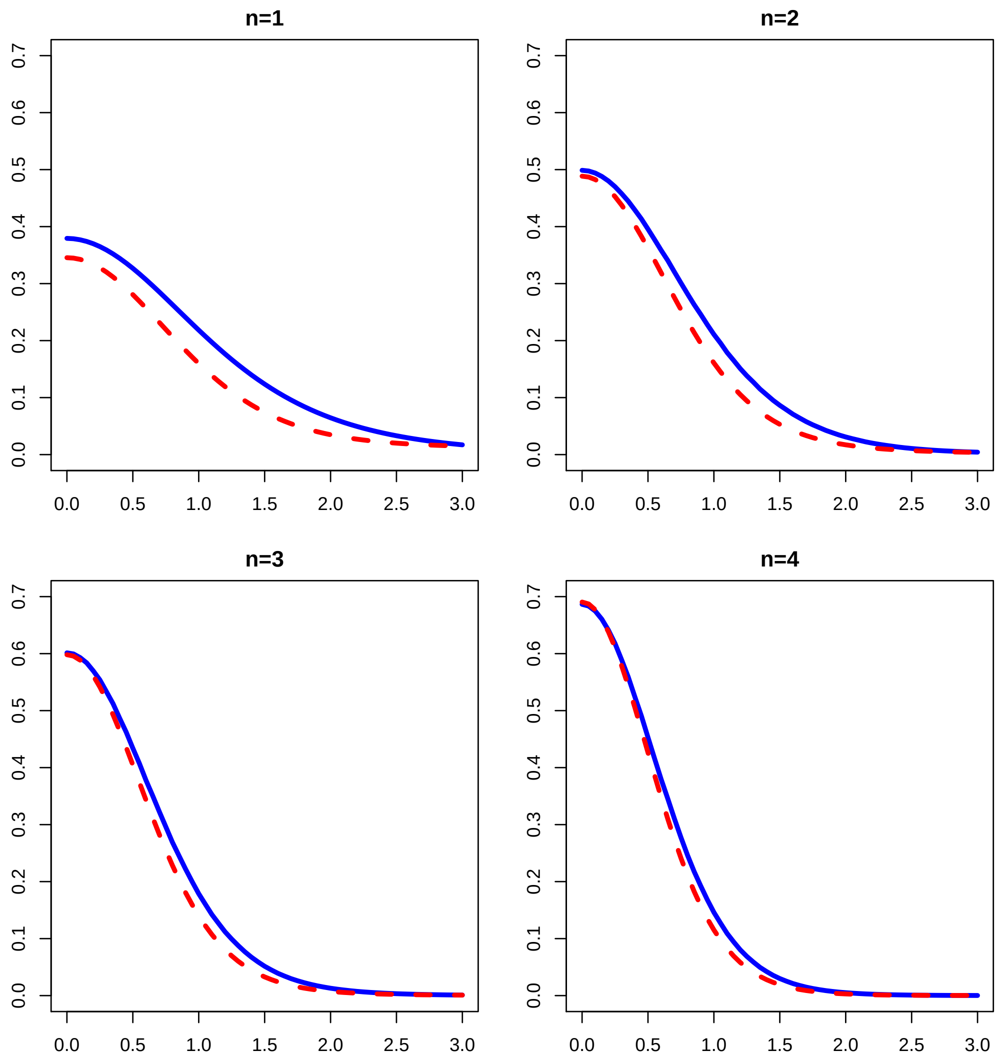


Figure 1.12: Comparison of exact density (blue line) of MLE and its saddlepoint approximation (red dashed line), Student t_5 distribution

where $\lambda(t)$ is the saddlepoint satisfying

$$\frac{\partial}{\partial\lambda}K_F(\lambda, t) = 0,$$

i.e.

$$\mathsf{E}_F[\psi(X_i, t)e^{\lambda^T\psi(X_i,t)}] = 0.$$

Moreover, in the case of simple hypothesis, the test statistic simplifies to

$$h(\hat{\theta}) = -K_F(\lambda(t), t).$$

If observations $X_1, \ldots, X_n$ are independent but not identically distributed, cumulant generating function is replaced by

$$K_F(\lambda, \theta) = \frac{1}{n}\sum_{i=1}^{n} K_{F^i}(\lambda, \theta),$$

where $K_{F^i}(\lambda, \theta) = \log \mathsf{E}_{F^i}[e^{\lambda^T\psi(X_i,\theta)}]$ and F^i is the distribution function of X_i. Under H_0

$$2nh(\hat{\theta}_1) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p_1}$$

with a relative error of order $O(n^{-1})$, for assumptions and the proof see Ma and Ronchetti (2011). This test is first-order equivalent to the three classical tests, but exhibits better second-order properties, i.e. has better small sample properties.

In case F is not known, empirical test based on empirical exponential likelihood may be used. Denote by F_0 the true cumulative distribution function of $X_1, \ldots, X_n$ and by $\hat{F}_0 = (w_1, \ldots, w_n)$ its empirical cdf that is closest to $(1/n, \ldots, 1/n)$ and is obtained by minimizing backward Kullback-Leibler divergence. The saddlepoint test statistic under $\hat{F}_0$ has the form

$$\hat{h}(\hat{\theta}) = \inf_{\theta_2}\sup_{\lambda}\left\{-K_0^w(\lambda, (\hat{\theta}_1, \theta_2))\right\},$$

where

$$K_0^w(\lambda, \theta) = \log\left(\sum_{i=1}^{n} w_i e^{\lambda^T\psi(x_i,\theta)}\right),$$

and weights w_i are computed as

$$w_i = e^{\mu(\theta^*)^T\psi(x_i,\theta^*)} / \sum_{j=1}^{n} e^{\mu(\theta^*)^T\psi(x_j,\theta^*)}, \tag{1.8}$$

where

$$\theta^* = (\theta_{1_0}, \theta_2^*)$$

$$\theta_2^* = \arg\min_{\theta_2}\{-\kappa(\lambda(\theta_{10}, \theta_2), (\theta_{10}, \theta_2))\} \tag{1.9}$$

$$\mu(\theta) = \arg\max_{\mu}\{-\kappa(\mu, \theta)\} \tag{1.10}$$

$$\kappa(\lambda, \theta) = \log\left(\frac{1}{n}\sum_{i=1}^{n} e^{\mu(\theta)^T\psi(x_i,\theta)}\right). \tag{1.11}$$

When $n \to \infty$, the p-value satisfies

$$P_{H_0}\{2n\hat{h}(\hat{\theta}_1) \geq 2n\hat{h}(\hat{\theta}_{1obs})\} = \{1 - Q_{p_1}(2n\hat{h}(\hat{\theta}_{1obs}))\{1 + O_P(n^{-1})\}\},$$

where Q_{p_1} denotes cdf of the χ^2 distribution with p_1 degrees of freedom. The assumptions and the proof can be found in (Ma and Ronchetti, 2011).

Chapter 2

Saddlepoint approximations of the density in quantile regression

In this chapter the distribution of regression quantiles will be studied. Quantile regression enables us to study the conditional quantiles of a dependent random variable and therefore offers a more complex view at its behavior. By appropriate choice of quantile we can study not only conditional median but also behavior in tails and get an idea about variability of a studied random variable. Although finite-sample distribution of regression quantiles has been already derived by various authors, its formula is not suitable for computing. Distribution of a simple quantile will be also studied as a special case of regression quantile.

2.1 Regression quantiles

Quantile regression was introduced in 1978 in a paper Koenker and Bassett (1978) as an alternative to nonrobust least squares estimators for the linear model. Efficiency of these estimators is comparable to least squares estimator (LSE) under normality, whilst it outperforms LSE in models with non-Gaussian errors. Instead of modelling the conditional expectation of the response given the covariates, models the α quantiles of the conditional distribution and provides a richer information on the underlying relationship between the response and the covariates. From the original formulation for the standard regression model, many extensions have been provided, including generalized linear models, survival data, autoregressive models, penalized methods, and nonparametric regression. Moreover, many applications in various fields ranging from economics, finance to biology and ecology have been developed. An excellent overview on theoretical, computational, and applied aspects is given in the book Koenker (2005).

Definition 1. *(Koenker and Bassett, 1978) Let us consider a linear model*

$$Y_i = \mathbf{X_i^T}\boldsymbol{\beta} + e_i, \tag{2.1}$$

where $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$*,* $\mathbf{X_i} = (1, X_{i1}, \ldots, X_{ip})$*, and* e_i *are i.i.d. with distribution function* F*. The* α*th regression quantile,* $0 < \alpha < 1$ *is defined as any solution to the* *minimization problem*

$$\min_{\mathbf{b}\in\mathbb{R}^{p+1}}\left[\sum_{i\in\{i:y_i\geq x_i^T\mathbf{b}\}}\alpha|y_i-x_i^T\mathbf{b}|+\sum_{i\in\{i:y_i<x_i^T\mathbf{b}\}}(1-\alpha)|y_i-x_i^T\mathbf{b}|\right]$$

Since its influence function (Hampel (1974), Hampel et al. (1986)) is proportional to the score function, it is bounded and the regression quantile estimator is robust against small deviations from the underlying error distribution and against possible outlying observations in the response variable.

Remark 2. *The minimization problem in the definition of regression quantile is equivalent to the following optimization problem, which solution are the so-called regression rank scores:*

$$\begin{aligned}&\max \mathbf{Y}^T\hat{a}\\ s.t.\ &\mathbf{X}^T\hat{a}=(1-\alpha)\mathbf{X}^T\mathbf{1_n}\\ &\hat{a}\in[0,1]^n.\end{aligned}$$

The asymptotic distribution of the regression quantile was derived by Koenker and Bassett (1978).

Theorem 2. *(Asymptotic distribution of regression quantiles) Let the following conditions be satisfied*

(A1) The distribution functions $\{F_i\}$ *are absolutely continuous, with continuous densities* $\{f_i\}$ *uniformly bounded away from* 0 *and* ∞ *at the points* $\xi_i(\alpha)=F_i^{-1}(\alpha)$, $i=1,2,\ldots$

(A2) There exist positive definite matrices $\mathbf{D}_0$ *and* $\mathbf{D}_1(\alpha)$ *such that*

- *(i)* $\lim_{n\to\infty}\frac{1}{n}\mathbf{X}^T\mathbf{X}=\mathbf{D}_0$,
- *(ii)* $\lim\frac{1}{n}\sum_{i=1}^n f_i(\xi_i(\alpha))\mathbf{x}_i\mathbf{x_i^T}=\mathbf{D}_1(\alpha)$,
- *(iii)* $\max_{i=1,\ldots,n}||\mathbf{x}_i||/\sqrt{n}\to 0$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_\alpha)\xrightarrow[n\to\infty]{\mathcal{D}}N(0,\alpha(1-\alpha)D_1^{-1}D_0D_1^{-1})$$

and in the iid error model

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta})\xrightarrow[n\to\infty]{\mathcal{D}}N\left(0,\frac{\alpha(1-\alpha)}{f_i^2(\xi_i(\alpha))}\mathbf{D}_0^{-1}\right).$$

Finite-sample distribution of regression quantiles was already derived in seminal paper Koenker and Bassett (1978), although the result is not suitable for computing. The density of a regression quantile for a finite number of observations was later also derived by (Jurečková, 2010).

Theorem 3. *Consider the linear regression model with deterministic regression matrix* $\mathbf{X}$ *of rank* p*, with the first column equal to* $\mathbf{1_n}$*. Assume that the errors* $e_1, \ldots, e_n$ *are i.i.d. with absolutely continuous distribution function* F *and with density* f*, absolutely continuous and positive for* $z \in (a, b)$ *where* $a = \inf\{z : F(z) > 0\}$ *and* $b = \sup\{z : F(z) < 1\}$*. Then the* α*th regression quantile* $\hat{\boldsymbol{\beta}}_\alpha$*,* $0 < \alpha < 1$ *has the density*

$$g(\mathbf{b}; \alpha) = \sum_{a \in \mathbf{A}_n(\alpha)} \prod_{i=1}^{n} \left[\left(F(\mathbf{x_i^T}[\mathbf{b} - \boldsymbol{\beta}])\right)^{\mathrm{I}[a_i=0]} \left(1 - F(\mathbf{x_i^T}[\mathbf{b} - \boldsymbol{\beta}])\right)^{\mathrm{I}[a_i=1]} \times \left(f(\mathbf{x_i^T}[\mathbf{b} - \boldsymbol{\beta}])\right)^{\mathrm{I}[0<a_i<1]} \right], \quad \mathbf{b} \in \mathbb{R}^{p+1},$$

where

$$\mathbf{A}_n(\alpha) = \Big\{ \mathbf{a} : 0 \le a_i \le 1, i = 1, \ldots, n, \sum_{i=1}^{n} \mathbf{x}_i a_i \,\mathrm{I}[a_i > 0] = (1 - \alpha) \sum_{i=1}^{n} \mathbf{x}_i \text{ and } 0 < a_{i_j} < 1, j = 1, \ldots, p \text{ for } 1 \le i_1, \ldots, i_p \le n \text{ such that } \mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_p} \text{ is a basis of } \mathbb{R}^p \Big\}.$$

While working with regression quantiles, a problem of estimating the quantity $q(\alpha) = \frac{1}{f(F^{-1}(\alpha))}$ often arises (the so called quantile density function), see the asymptotic distribution in Theorem 2. One of the methods how to solve this was proposed in Dodge and Jurečková (2000) where a histogram-type estimator was suggested

$$H_n(\alpha) = \frac{1}{2\nu_n} \left[\hat{\boldsymbol{\beta}}_{\alpha 1}(\alpha + \nu_n) - \hat{\boldsymbol{\beta}}_{\alpha 1}(\alpha - \nu_n) \right],$$

where $\hat{\boldsymbol{\beta}}_{\alpha 1}$ denotes the first component of a regression quantile and the recommended choice of bandwidth ν_n is

$$\nu_n = \left(\frac{q(\alpha)}{q''(\alpha)} \right)^{2/5} \left(\frac{9}{2} \right)^{1/5} n^{-1/5}.$$

The population counterpart of the α-regression quantile $\hat{\boldsymbol{\beta}}_\alpha$ is the α-population regression quantile $(\beta_0 + F^{-1}(\alpha), \beta_1, \ldots, \beta_p)$.

Even though regression quantile is an M-estimator, due to the form of ρ-function, the assumptions on smoothness of its derivative required by saddlepoint techniques are usually not satisfied. Let us mention a useful result by (Ruppert and Carroll, 1980).

Lemma 1. *Ruppert and Carroll (1980) Let* $r_1, \ldots, r_n$ *be the residuals from preliminary estimate* $\boldsymbol{\beta}_0$*, suppose* $0 < \alpha < 1$*, and let* μ_n *be a sequence of solutions to*

$$\sum_{i=1}^{n} \rho_\alpha(r_i - \mu_n) = min.$$

Then

$$n^{-1/2} \sum_{i=1}^{n} \psi_\alpha(r_i - \mu_n) \to 0 \ \ a.s.$$

In addition, the sequence of solutions $\hat{\boldsymbol{\beta}}_\alpha$ *satisfies*

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{x}_i\psi_\alpha(y_i-\mathbf{x}_i\hat{\boldsymbol{\beta}}_\alpha)\xrightarrow[n\to\infty]{}0\ a.s.$$

2.2 Saddlepoint approximation to the density of regression quantile

Saddlepoint approximations to the density of M-estimators have proved to be very accurate. Unfortunately, one of their drawbacks is a need of numerical methods when computing saddlepoint and explicit solution is only rarely available. Because the ρ-function and its directional derivatives for regression quantiles is very simple, it allows us to write not only the saddlepoint, but the approximation to the density in explicit form.

In the following computations we will use

$$\psi_\alpha(Y_i,\mathbf{b})=(\alpha-\mathrm{I}[Y_i-\mathbf{X}_i^T\mathbf{b}<0])\mathbf{X}_i.$$

Notice, that $\psi_\alpha(\cdot)$ is bounded (this implies the existence of moment generating function of ψ-function) and the derivative does not exist in one point.

2.2.1 Parametric case

To carry out the computations in this case, we assume for convenience that $(Y_i,\mathbf{x}_i)$ are independent identically distributed with density $g(y_i-\mathbf{x}_i^T\boldsymbol{\beta})k(\mathbf{x}_i)$, where $k(\cdot)$ is the density of $\mathbf{x}_i$.

Theorem 4. *Consider model* (2.1) *and assume that conditions (S1) – (S5) are satisfied. Then the asymptotic expansion for the density of a regression quantile is*

$$f_n(\mathbf{b})=\left(\frac{n}{2\pi}\right)^{1/2}c^{-n}(\mathbf{b})A(\mathbf{b})/\sigma(\mathbf{b})[1+O(n^{-1})],$$

where

$$\begin{aligned}c^{-1}(\mathbf{b})=&\frac{\alpha^{\alpha-1}}{(1-\alpha)^\alpha}\int\frac{\left(1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)^{\alpha-1}}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^{\alpha}}\left[G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^2+\alpha\right.\\&\left.-2\alpha G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right]k(\mathbf{x}_i)d\mathbf{x}_i\\A(\mathbf{b})=&c(\mathbf{b})\int\left\{(1-\alpha)\mathbf{x}_i\left[\frac{1-\alpha}{\alpha}\frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right]^{\alpha-1}g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)+\right.\\&\left.+\alpha\mathbf{x}_i\left[\frac{1-\alpha}{\alpha}\frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right]^{\alpha}g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right\}\mathbf{x}_ik(\mathbf{x}_i)d\mathbf{x}_i\end{aligned}$$

and

$$\sigma^2(\mathbf{b}) = c(\mathbf{b}) \int \frac{\left[(1-\alpha)G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right]^{\alpha}}{\left[\left(1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)\alpha\right]^{\alpha-1}} k(\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i^T d\mathbf{x}_i.$$

Remark 3. *The saddlepoint approximation to the density of regression quantile equals*

$$g_n(\mathbf{b}) = \left(\frac{n}{2\pi}\right)^{1/2} c^{-n}(\mathbf{b})A(\mathbf{b})/\sigma(\mathbf{b}).$$

Proof. The saddlepoint approximation to the density of regression quantile will be derived using Theorem 1, in particular we use the formula

$$f_n(\mathbf{b}) = \left(\frac{n}{2\pi}\right)^{1/2} c^{-n}(\mathbf{b})A(\mathbf{b})/\sigma(\mathbf{b})[1+O(n^{-1})],$$

where $\gamma(\mathbf{b})$ is a solution of the equation

$$\int \psi(y,\mathbf{b}) \exp\{\gamma\psi(y,\mathbf{b})\} f(y)dy = 0$$

and

$$c^{-1}(\mathbf{b}) = \int \exp\{\gamma\psi(y,\mathbf{b})\} f(y)dy$$
$$\sigma^2(\mathbf{b}) = \mathsf{E}_{\mathbf{b}}\psi^2(y,\mathbf{b})$$
$$A(\mathbf{b}) = \mathsf{E}_{\mathbf{b}}D\psi(y,\mathbf{b}),$$

where $\mathsf{E}_{\mathbf{b}}$ is expected value with respect to the conjugate density

$$h_{\mathbf{b}}(y) = c(\mathbf{b})exp\{\gamma(\mathbf{b})\psi(y,\mathbf{b})\}f(y).$$

The saddlepoint $\gamma(\mathbf{b})$ satisfies the following equation

$$\int (\alpha - \mathrm{I}[y_i - \mathbf{x}_i^T\mathbf{b} < 0])\mathbf{x}_i e^{\gamma(\alpha - \mathrm{I}[y_i - \mathbf{x}_i^T\mathbf{b}<0])\mathbf{x}_i} \frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i)d\mathbf{x}_i dy_i = 0.$$

Using the form of ψ-function we can rewrite this equation in the form

$$\int\int \alpha\mathbf{x}_i e^{\gamma\alpha\mathbf{x}_i} e^{-\gamma\,\mathrm{I}[y_i<\mathbf{x}_i^T\mathbf{b}]\mathbf{x}_i} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i)d\mathbf{x}_i dy_i =$$

$$\int\int \mathrm{I}[y_i < \mathbf{x}_i^T\mathbf{b}]\mathbf{x}_i e^{\gamma\alpha\mathbf{x}_i} e^{-\gamma\,\mathrm{I}[y_i<\mathbf{x}_i^T\mathbf{b}]\mathbf{x}_i} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i)d\mathbf{x}_i dy_i$$

and using the indicator functions it simplifies to

$$\alpha \int \left(\int_{-\infty}^{\mathbf{x}_i^T\mathbf{b}} e^{-\gamma\mathbf{x}_i} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i + \int_{\mathbf{x}_i^T\mathbf{b}}^{\infty} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i \right) \mathbf{x}_i e^{\gamma\alpha\mathbf{x}_i} k(\mathbf{x}_i)d\mathbf{x}_i =$$

$$\int \left(\int_{-\infty}^{\mathbf{x}_i^T\mathbf{b}} \exp\{-\gamma \mathbf{x}_i\} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i \right) e^{\gamma\alpha \mathbf{x}_i} \mathbf{x}_i k(\mathbf{x}_i) d\mathbf{x}_i$$

$$\begin{aligned}(1-\alpha)\int_{-\infty}^{\mathbf{x}_i^T\mathbf{b}} e^{-\gamma \mathbf{x}_i} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i =& \alpha \int_{\mathbf{x}_i^T\mathbf{b}}^{\infty} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i \\ e^{-\gamma(\mathbf{b})\mathbf{x}_i} =& \frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\end{aligned}$$

It is not possible to find explicit formula for saddlepoint $\gamma(\mathbf{b})$, in the following computations it is sufficient to consider the following function of saddlepoint

$$-\gamma(\mathbf{b})\mathbf{x}_i = \log \frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}.$$

Based on the derived saddlepoint, we can compute the following expectations w.r.t. conjugate density (see Theorem 1.1)

$$\begin{aligned}c^{-1}(\mathbf{b}) =& \int \exp\{\gamma\psi(y_i, \mathbf{b})\} \frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i) dy_i d\mathbf{x}_i \\ =& \int\int \left(\frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right)^{\alpha - \mathrm{I}[y_i < \mathbf{b}]} \frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i) dy_i d\mathbf{x}_i \\ =& \frac{1}{\sigma}\int \left(\int_{-\infty}^{\mathbf{x}_i^T\mathbf{b}} \left(\frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right)^{\alpha-1} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i + \right. \\ & \left. + \int_{\mathbf{x}_i^T\mathbf{b}}^{\infty} \left(\frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right)^{\alpha} g\left(\frac{y_i - \mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right) dy_i \right) k(\mathbf{x}_i) d\mathbf{x}_i \\ =& \int \left(\frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right)^{\alpha-1} G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) + \\ & + \left(\frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}\right)^{\alpha} (1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)) k(\mathbf{x}_i) d\mathbf{x}_i \\ =& \frac{\alpha^{\alpha-1}}{(1-\alpha)^{\alpha}} \int \frac{\left(1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)^{\alpha-1}}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^{\alpha}} \left[G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^2 + \alpha \right. \\ & \left. - 2\alpha G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) \right] k(\mathbf{x}_i) d\mathbf{x}_i\end{aligned}$$

$$\begin{aligned}
\sigma^2(\mathbf{b}) =& c(\mathbf{b}) \int \int (\alpha - \mathrm{I}[y_i < \mathbf{x}_i^T \mathbf{b}])^2 \mathbf{x}_i \mathbf{x}_i^T \frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x}_i) * \\
& * e^{-\log \frac{\alpha}{1-\alpha} \frac{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} (\alpha - \mathrm{I}[y_i < \mathbf{x}_i^T \mathbf{b}])} d\mathbf{x}_i \\
=& c(\mathbf{b}) \int \left[(\alpha-1)^2 \left(\frac{(1-\alpha) G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{\alpha\left(1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)} \right)^{\alpha-1} G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) + \right. \\
& \left. + \alpha^2 \left(\frac{(1-\alpha) G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{\alpha(1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right))} \right)^{\alpha} \left(1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right) \right] k(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T d\mathbf{x}_i \\
=& c(\mathbf{b}) \int \frac{\left[(1-\alpha) G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right]^{\alpha}}{\left[\left(1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)\alpha\right]^{\alpha-1}} k(\mathbf{x}_i) \mathbf{x}_i \mathbf{x}_i^T d\mathbf{x}_i
\end{aligned}$$

$$\begin{aligned}
A(\mathbf{b}) =& \mathsf{E}_{\mathbf{b}} D\psi(y_i, \mathbf{b}) \\
=& c(\mathbf{b}) \int \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{(\alpha - \mathrm{I}[y_i < \mathbf{x}_i^T \mathbf{b}])} \frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \\
& \times k(\mathbf{x}_i) D\psi(y_i, \mathbf{b}) dy_i d\mathbf{x}_i \\
=& c(\mathbf{b}) \int \left\{ \left[\psi(y_i, \mathbf{b}) g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{(\alpha - \mathrm{I}[y_i < \mathbf{x}_i^T \mathbf{b}])} \right]_{-\infty}^{\infty} - \right. \\
& \left. - \int_{-\infty}^{\infty} \psi(y_i, \mathbf{b}) \left(\frac{1}{\sigma} g\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{(\alpha - \mathrm{I}[y_i < \mathbf{x}_i^T \mathbf{b}])} \right)' dy_i \right\} \\
& \times \mathbf{x}_i k(\mathbf{x}_i) d\mathbf{x}_i \\
=& c(\mathbf{b}) \int \left\{ \int_{-\infty}^{\mathbf{x}_i^T \mathbf{b}} (1-\alpha) \mathbf{x}_i \frac{1}{\sigma} g'\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{(\alpha-1)} - \right. \\
& \left. - \int_{\mathbf{x}_i^T \mathbf{b}}^{\infty} \alpha \mathbf{x}_i \frac{1}{\sigma} g'\left(\frac{y_i - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}\right) \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{\alpha} \right\} \mathbf{x}_i k(\mathbf{x}_i) d\mathbf{x}_i
\end{aligned}$$

$$
\begin{aligned}
=&c(\mathbf{b}) \int \left\{ (1-\alpha)\mathbf{x}_i \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{\alpha-1} g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) + \right. \\
&\left. +\alpha\mathbf{x}_i \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{\alpha} g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) \right\} \mathbf{x}_i k(\mathbf{x}_i) d\mathbf{x}_i
\end{aligned}
$$

□

Remark 4. *In case the matrix* $\mathbf{X}$ *is known, the saddlepoint approximation to the density of regression quantile equals*

$$
f_n(\mathbf{b}) = \left(\frac{n}{2\pi}\right)^{1/2} c^{-n}(\mathbf{b}) A(\mathbf{b})/\sigma(\mathbf{b})[1+O(n^{-1})],
$$

where

$$
\begin{aligned}
c^{-1}(\mathbf{b}) =& \frac{\alpha^{\alpha-1}}{(1-\alpha)^{\alpha}} \sum_{i=1}^{n} \frac{\left(1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)^{\alpha-1}}{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^{\alpha}} \left[G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)^2 + \alpha \right. \\
& \left. -2\alpha G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) \right] \\
A(\mathbf{b}) =& c(\mathbf{b}) \sum_{i=1}^{n} \left\{ (1-\alpha)\mathbf{x}_i \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{\alpha-1} g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) + \right. \\
& \left. +\alpha\mathbf{x}_i \left[\frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)}{1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)} \right]^{\alpha} g\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right) \right\} \mathbf{x}_i \\
\sigma^2(\mathbf{b}) =& c(\mathbf{b}) \sum_{i=1}^{n} \frac{\left[(1-\alpha)G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right]^{\alpha}}{\left[\left(1-G\left(\frac{\mathbf{x}_i^T(\mathbf{b}-\boldsymbol{\beta})}{\sigma}\right)\right)\alpha\right]^{\alpha-1}} \mathbf{x}_i \mathbf{x}_i^T.
\end{aligned}
$$

As a special case, we may compute the saddlepoint approximation to the density of a sample quantile.

Corrollary 1. *Assume the conditions (S1) – (S5) are satisfied. The saddlepoint approximation to the density of a sample* α*-quantile equals*

$$
f_n(b) = -\sqrt{\frac{n}{2\pi}} \left(\frac{1-\alpha}{1-G(b)}\right)^{n(\alpha-1)} \left(\frac{G(b)}{\alpha}\right)^{n\alpha} \frac{g(b)}{\log\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)} \frac{\alpha - G(b)}{G(b)(1-G(b))} \sqrt{\alpha(1-\alpha)}.
$$

Proof. The computations of saddlepoint $\gamma(b)$, $c^{-1}(b)$ and $\sigma^2(b)$ are straightforward and we get

$$\begin{aligned}\gamma(b) =& \log \frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\\ c^{-1}(b) =& \left(\frac{1-\alpha}{1-G(b)}\right)^{\alpha-1}\left(\frac{G(b)}{\alpha}\right)^{\alpha}\\ \sigma^2(b) =& \alpha(1-\alpha),\end{aligned}$$

where the conjugate density equals

$$h_b(y) = \left(\frac{1-\alpha}{1-G(b)}\right)^{1-\alpha}\left(\frac{\alpha}{G(b)}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha-\mathrm{I}[y<b]} g(y).$$

When computing $A(b)$, in order to deal with the derivative of ψ-function, the expectation has to be computed using integration by parts:

$$\begin{aligned}A(b) =& \mathsf{E}_b D\psi(y,b)\\ =& \left(\frac{1-\alpha}{1-G(b)}\right)^{1-\alpha-1}\left(\frac{\alpha}{G(b)}\right)^{\alpha}\int_{-\infty}^{\infty}\frac{\partial\psi(y,b)}{\partial y}\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha-\mathrm{I}[y<b]} g(y)dy\\ =& -\frac{1}{\log\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)}\left(\frac{1-\alpha}{1-G(b)}\right)^{1-\alpha}\left(\frac{\alpha}{G(b)}\right)^{\alpha}\left\{\left[\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha-\mathrm{I}[y<b]} g(y)\right]_{-\infty}^{\infty}\right.\\ & \left.-\int_{-\infty}^{\infty}\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha-\mathrm{I}[y<b]} g'(y)dy\right\}\\ =& -\frac{1}{\log\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)}\left(\frac{1-\alpha}{1-G(b)}\right)^{1-\alpha}\left(\frac{\alpha}{G(b)}\right)^{\alpha}\left\{\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha-1} g(b)\right.\\ & \left.-\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)^{\alpha} g(b)\right\}\\ =& -\frac{g(b)}{\log\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)}\frac{\alpha-G(b)}{G(b)(1-G(b))}.\end{aligned}$$

Thus the saddlepoint approximation to the density of a sample quantile equals

$$-\sqrt{\frac{n}{2\pi}}\left(\frac{1-\alpha}{1-G(b)}\right)^{n(\alpha-1)}\left(\frac{G(b)}{\alpha}\right)^{n\alpha}\frac{g(b)}{\log\left(\frac{1-\alpha}{\alpha}\frac{G(b)}{1-G(b)}\right)}\frac{\alpha-G(b)}{G(b)(1-G(b))}\sqrt{\alpha(1-\alpha)}.$$

□

Because the formula for exact density of order statistic is available, we might compare it to the derived approximation. In the following numerical study we considered sample quantiles of standard normal distribution and exponential distribution with parameter $\lambda = 2$ for different sample sizes ($n = 11, 21, 31, 41$) and different values of α ($\alpha = 0.1, 0.3, 0.5$). The exact density is plotted by red lines, the approximation by black lines. As expected, these two differ mostly for $\alpha = 0.1$. When increasing α and sample size, the saddlepoint approximation proves to be very precise.

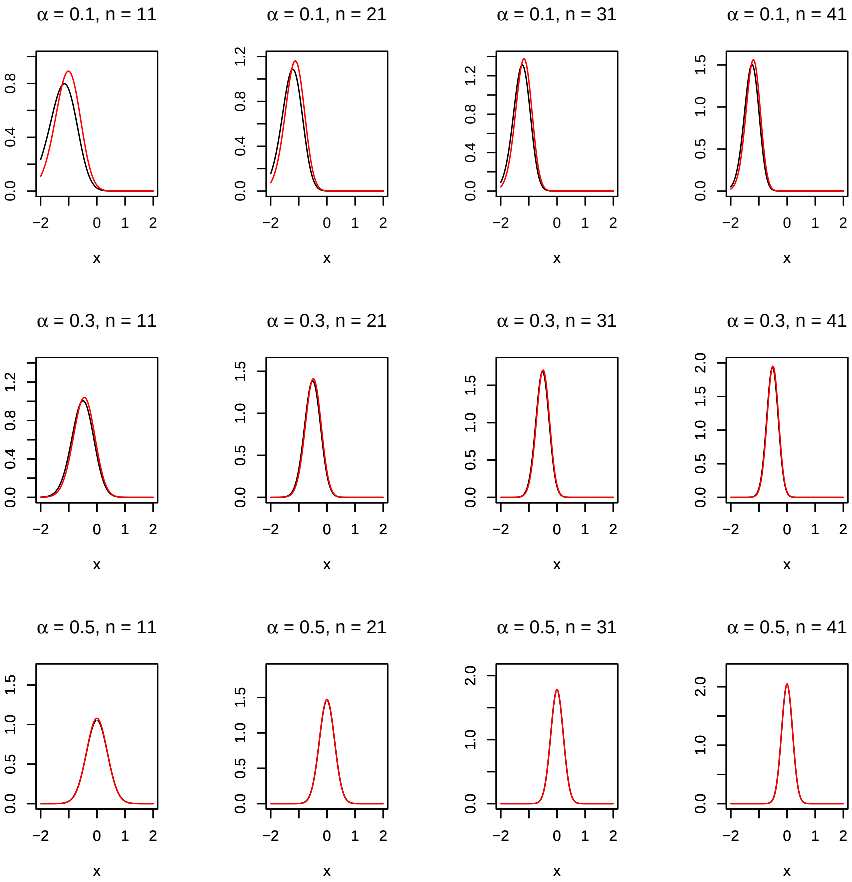


Figure 2.1: Density of a sample quantile, comparison of exact density (red) and its saddlepoint approximation (black), normal distribution

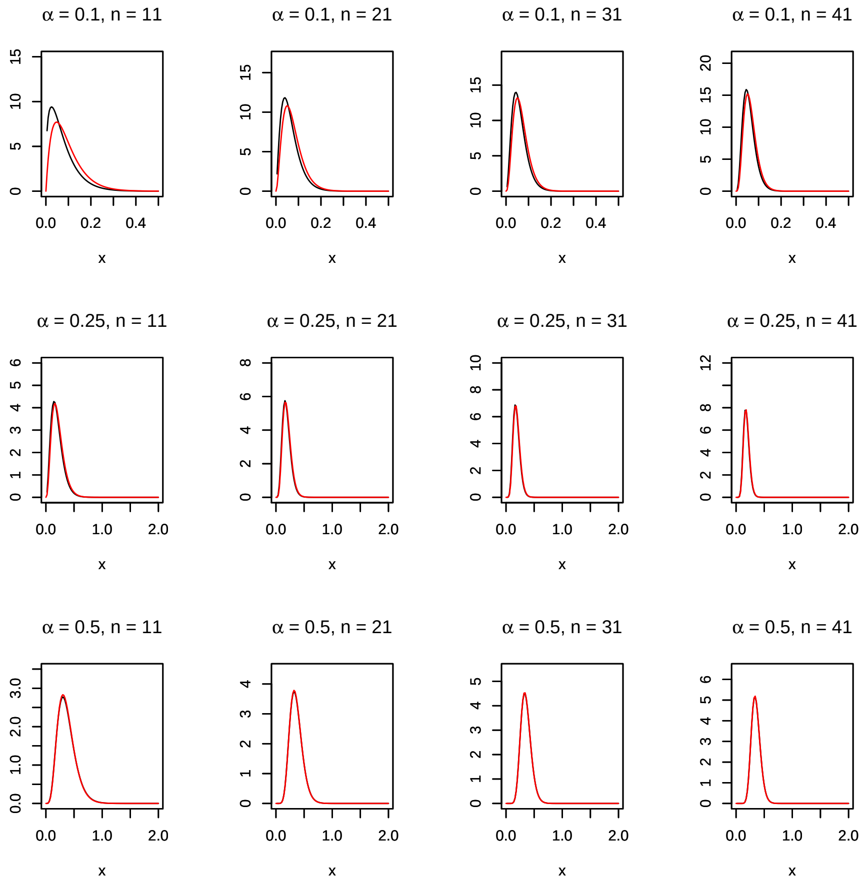


Figure 2.2: Density of a sample quantile, comparison of exact density (red) and its saddlepoint approximation (black), exponential distribution

2.2.2 Nonparametric case

In case the distribution of errors in regression model is not known, it is possible to use the so-called empirical saddlepoint approximation. Let us consider model

$$Y_i = x_i^T \boldsymbol{\beta} + e_i, \qquad i = 1, \dots, n,$$

where $e_1, \dots, e_n$ are i.i.d. with distribution function F.

α-regression quantile $\boldsymbol{\beta}_\alpha$ is an approximate solution of

$$\sum_{i=1}^{n} (\alpha - I[y_i - x_i^T \boldsymbol{\beta} < 0]) x_i = 0.$$

Let us denote

$$\begin{aligned} I_{ij}(\mathbf{b}) =& I[r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b}) < 0], \\ F_n^j(\mathbf{b}) =& \frac{1}{n} \sum_{i=1}^{n} I_{ij}(\mathbf{b}). \end{aligned}$$

Empirical saddlepoint density approximations were developed in Ronchetti and Welsh (1994). In case of the regression model, the derivation of the approximation to the density is a bit more complicated than in i.e. location model, as it is necessary to consider the empirical distribution function of residuals $r_i = y_i - \mathbf{x_i^T} \hat{\boldsymbol{\beta}}_\alpha$, $i = 1, \dots, n$. The computations remind of those in the parametric case where the integrals are usually replaced by the sums.

Theorem 5. *Consider model* (2.1) *and assume that conditions (a) – (e) in Ronchetti and Welsh (1994) are satisfied. Then the saddlepoint approximation to the density of regression quantile in case the distribution of* $e_1, \dots, e_n$ *is unspecified, equals*

$$\hat{g}_n(\mathbf{b}) = \left(\frac{n}{2\pi}\right)^{s/2} \left|\hat{\Sigma}(\mathbf{b})\right|^{-1/2} \left|\hat{A}(\mathbf{b})\right| \exp\left\{\sum_{j=1}^{n} \hat{K}^j(\mathbf{b})\right\}$$

where

$$\begin{aligned} \hat{K}^j =& \log\left(\frac{1}{n}\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{\alpha} \sum_{i=1}^{n}\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{I_{ij}(\mathbf{b})}\right) \\ \hat{\Sigma}(\mathbf{b}) =& \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}} \left\{\sum_{i=1}^{n}(\alpha - I_{ij}(\mathbf{b}))^2 \left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{ij}(\mathbf{b})}\right\} x_j x_j^T \\ \hat{A}(\mathbf{b}) =& \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}} \sum_{i=1}^{n}\left(\frac{\partial}{\partial \mathbf{b}}(\alpha - I_{ij}(\mathbf{b}))x_j\right)\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{-I_{ij}(\mathbf{b})} \end{aligned}$$

Proof. The saddlepoint approximation to the density (see Ronchetti and Welsh (1994)) equals

$$\hat{g}_n(\mathbf{b}) = \left(\frac{n}{2\pi}\right)^{s/2} \left|\hat{\Sigma}(\mathbf{b})\right|^{-1/2} \left|\hat{A}(\mathbf{b})\right| \exp\left\{\sum_{j=1}^{n} \hat{K}^j(\mathbf{b})\right\}$$

where $\hat{K}$, $\hat{A}$ and $\hat{\Sigma}$ can be considered as the empirical counterparts to the terms in the saddlepoint approximation in parametric case. The saddlepoint $\hat{\gamma}(\mathbf{b})$ satisfies

$$\sum_{j=1}^{n}\sum_{i=1}^{n}(\alpha - I[r_i + x_j^T(\hat{\boldsymbol{\beta}} - \mathbf{b}) < 0])x_j \exp\left\{\hat{\gamma}(\mathbf{b})^T(\alpha - I[r_i + x_j^T(\hat{\boldsymbol{\beta}} - \mathbf{b}) < 0])x_j\right\} = 0.$$

Hence we again have the formula for the function of $\hat{\gamma}(\mathbf{b})$

$$\hat{\gamma}(\mathbf{b})^T x_j = \log\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})},$$

where

$$\begin{aligned} I_{ij}(\mathbf{b}) =& I[r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b}) < 0] \\ F_n^j(\mathbf{b}) =& \frac{1}{n}\sum_{i=1}^{n} I_{ij}(\mathbf{b}). \end{aligned}$$

The empirical cumulant generating function equals

$$\begin{aligned} \hat{K}^j(\mathbf{b}) =& \log\left(\frac{1}{n}\sum_{i=1}^{n}\exp\left\{\hat{\gamma}(\mathbf{b})^T(\alpha - I_{ij}(\mathbf{b}))x_j\right\}\right) \\ =& \log\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{\alpha - I_{ij}(\mathbf{b})}\right) \\ =& \log\left(\frac{1}{n}\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{\alpha}\sum_{i=1}^{n}\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{I_{ij}(\mathbf{b})}\right) \end{aligned}$$

The empirical counterpart to matrix Σ equals

$$\begin{aligned} \hat{\Sigma}(\mathbf{b}) =& \frac{1}{n}\sum_{j=1}^{n}\exp\left\{-\hat{K}^j(\mathbf{b})\right\}\frac{1}{n}\sum_{i=1}^{n}(\alpha - I_{ij}(\mathbf{b}))^2 x_j x_j^T \exp\left\{(\alpha - I_{ij}(\mathbf{b}))\hat{\gamma}(\mathbf{b})^T x_j\right\} \\ =& \frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} n\frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{\alpha}}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}}(\alpha - I_{ij}(\mathbf{b}))^2 x_j x_j^T\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{\alpha - I_{ij}(\mathbf{b})} \\ =& \frac{1}{n}\sum_{j=1}^{n}\frac{1}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}}\left\{\sum_{i=1}^{n}(\alpha - I_{ij}(\mathbf{b}))^2\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{ij}(\mathbf{b})}\right\}x_j x_j^T. \end{aligned}$$

Finally, the expectation of derivative of ψ-function is in nonparametric case replaced by

$$\begin{aligned}
\hat{A}(\mathbf{b}) =& \frac{1}{n^2}\sum_{j=1}^{n} \exp\left\{-\hat{K}^j(\mathbf{b})\right\} \sum_{i=1}^{n} \left(\frac{\partial}{\partial \mathbf{b}}(\alpha - I_{ij}(\mathbf{b}))x_j\right) \exp\left\{\hat{\gamma}(\mathbf{b})^T(\alpha - I_ij(\mathbf{b}))x_j\right\} \\
=& \frac{1}{n^2}\sum_{j=1}^{n} n\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{\alpha} \frac{1}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}} \sum_{i=1}^{n}\left(\frac{\partial}{\partial \mathbf{b}}(\alpha - I_{ij}(\mathbf{b}))x_j\right) \times \\
& \left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{\alpha - I_{ij}(\mathbf{b})} \\
=& \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\sum_{l=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\mathbf{b})}{F_n^j(\mathbf{b})}\right)^{I_{lj}(\mathbf{b})}} \sum_{i=1}^{n}\left(\frac{\partial}{\partial \mathbf{b}}(\alpha - I_{ij}(\mathbf{b}))x_j\right)\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\mathbf{b})}{1-F_n^j(\mathbf{b})}\right)^{-I_{ij}(\mathbf{b})}
\end{aligned}$$

and this concludes the proof. □

Remark 5. *The derivative of indicator function $I_{ij}(\mathbf{b})$ might be approximated using the following approximation for the signum function*

$$\text{sign}(x) \approx \frac{x}{\sqrt{x^2 + \varepsilon^2}},$$

thus

$$\begin{aligned}
I_{ij}(b) =& \frac{1}{2}(1 - \text{sign}(r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b}))) \\
\approx& \frac{1}{2}\left(1 - \frac{r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b})}{\sqrt{(r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b}))^2 + \varepsilon^2}}\right)
\end{aligned}$$

and the derivative with respect to $\mathbf{b}$ *might be approximated by the following expression*

$$-\frac{1}{2}(\varepsilon^2 x_j^T)/(\varepsilon^2 + r_i + x_j^T(\hat{\boldsymbol{\beta}}_\alpha - \mathbf{b})^2)^{(3/2)}.$$

2.3 Density of averaged regression quantiles

The results derived for regression quantiles might be also used in derivation of the saddlepoint approximation to the density of averaged regression quantile. In this Section the procedure of estimating the density of averaged regression quantiles will be outlined. We will follow the development introduced in Fan and Field (1995).

Let us again assume that $(Y_i, \mathbf{x}_i)$ are independent identically distributed with density $g(y_i - \mathbf{x}_i^T\boldsymbol{\beta})k(\mathbf{x}_i)$, where $k(\cdot)$ is the density of $\mathbf{x}_i$. Suppose we want to compute the density of $\bar{x}\boldsymbol{\beta}$, where $\bar{\mathbf{x}} = (1, \bar{x}_1, \ldots, \bar{x}_p)^T = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ see Theorem 1 in Fan and Field (1995) for linear function $u(\mathbf{t}_0) = \bar{\mathbf{x}}\mathbf{t}_0$. Derivatives of function u w.r.t. to components of vector $\boldsymbol{\beta}$ equal

$$u^{(i)} = \frac{\partial u}{\partial \beta_i} = \bar{\mathbf{x}}_i, \qquad i = 1, \ldots, p+1.$$

Denote by $\mathbf{b}$ a solution of

$$\frac{1}{n}\sum_{i=1}^{n}\psi_l(Y_l,\mathbf{b})=0,$$

where $\boldsymbol{\psi}_l=(\alpha-\mathsf{I}[y_l-x_l^T\boldsymbol{\beta}<0])x_l$ and $\boldsymbol{\psi}_l=(\psi_{1l},\dots,\psi_{pl})$.

The conjugate density $h_{l,\mathbf{t}_0}$ to density of $(Y_l,\mathbf{X_l})$ has the form

$$h_{l,\mathbf{t}_0}(y_l,\mathbf{x}_l)=c_l(\mathbf{t}_0)\exp\left\{\sum_{j=1}^{p}\gamma_j\psi_{jl}(y_l,t_0)\right\}f_l(y_l,\mathbf{x}_l),$$

where $c_l(\mathbf{t}_0)$ is a normalizing constant so that $\int h_{l,\mathbf{t}_0}(y_l,\mathbf{x}_l)\,\mathrm{d}y_l\,\mathrm{d}\mathbf{x}_l=1$ and $\boldsymbol{\gamma}$ satisfies

$$\mathsf{E}_h\left\{\frac{1}{n}\sum_{l=1}^{n}\psi_l(Y_l,\mathbf{t}_0)\right\}=\mathbf{0}.$$

The saddlepoint $\boldsymbol{\gamma}$ satisfies

$$\int(\alpha-\mathsf{I}[y_l-\mathbf{x_l^T}\mathbf{t}_0<0])\mathbf{x_l}e^{\gamma^T(\alpha-\mathsf{I}[y_l-\mathbf{x_l^T}\mathbf{t}_0<0])\mathbf{x_l}}\frac{1}{\sigma}g\left(\frac{y_l-\mathbf{x_i^T}\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x_l})\,\mathrm{d}\mathbf{x_l}\,\mathrm{d}y_l=0$$

and following similar steps used in the previous parts of this chapter we get

$$-\gamma(\mathbf{t}_0)\mathbf{x_l}=\log\frac{\alpha}{1-\alpha}\frac{1-G\left(\frac{\mathbf{x_l^T}(\mathbf{t}_0-\boldsymbol{\beta})}{\sigma}\right)}{G\left(\frac{\mathbf{x_l^T}(\mathbf{t}_0-\boldsymbol{\beta})}{\sigma}\right)}.$$

Let us denote

$$\mathbf{A}(\mathbf{t}_0)=\mathsf{E}_h\left\{\frac{1}{n}\sum_{l=1}^{n}\psi_l^{(j)}(Y_l,\boldsymbol{\beta})|_{\boldsymbol{\beta}=\mathbf{t}_0}\right\}_{j=1,\dots,p},$$

where $h=\prod h_l$ and

$$\psi_l^{(j)}=\frac{\partial\psi_l}{\partial\boldsymbol{\beta}_j}$$

and compute the expectation w.r.t. the conjugate density

$$\begin{aligned}A(\mathbf{t}_0)=&\mathsf{E}_h\left\{\frac{1}{n}\sum_{l=1}^{n}\psi_l^{(j)}(Y_l,\boldsymbol{\beta})|_{\boldsymbol{\beta}=\mathbf{t}_0}\right\}_{j=1,\dots,p}\\
=&\int\int\frac{1}{n}\sum_{l=1}^{n}\psi_l^{(j)}(y_l,\boldsymbol{\beta})|_{\boldsymbol{\beta}=t_0}h_{l,t_0}\,\mathrm{d}\mathbf{x_l}\,\mathrm{d}y_l\\
=&-\frac{1}{\prod c_l(\mathbf{t}_0)}\int\int\frac{1}{n}\sum_{l=1}^{n}\frac{\partial\psi_l}{\partial y_l}|_{\boldsymbol{\beta}=t_0}\mathbf{x_l}\frac{1}{\sigma}g\left(\frac{y_l-\mathbf{x_l^T}\boldsymbol{\beta}}{\sigma}\right)\\
&\times k(\mathbf{x_l})\exp\left\{\sum_{j=1}^{p}\gamma_j\psi_{jl}(y_l,t_0)\right\}dx_l dy_l\end{aligned}$$

$$= -\frac{1}{\prod c_l(\mathbf{t_0})} \int\int \frac{1}{n}\sum_{l=1}^{n} \frac{\partial \psi_l}{\partial y_l}|_{\boldsymbol{\beta}=t_0} \mathbf{x_l} \frac{1}{\sigma} g\left(\frac{y_l - \mathbf{x_l^T}\boldsymbol{\beta}}{\sigma}\right) k(\mathbf{x_l})$$
$$\times \left[\frac{(1-\alpha)G_l}{\alpha(1-G_l)}\right]^{\alpha - \mathbb{I}[y_l - \mathbf{x_l^T}\mathbf{t}_0 < 0]} \mathrm{d}\mathbf{x_l}\,\mathrm{d}y_l$$

where

$$G_l = G\left(\frac{\mathbf{x_l^T}(\mathbf{t}_0 - \boldsymbol{\beta})}{\sigma}\right).$$

As a next step, the inverse of matrix A has to be computed numerically

$$B(\mathbf{t}_0) = -A(\mathbf{t}_0)^{-1}$$

and finally, the density of the averaged regression quantile is equal to

$$\left(\prod_{l=1}^{n} c_l(t_0)\right)^{-1} \sqrt{\frac{1}{2\pi\sigma^2_{G|h}}},$$

where $\sigma^2_{G|h}$ is the variance of

$$G = u_0 + \frac{1}{n}\sum_{l=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} u^k(\mathbf{t}_0) B_{kj}\psi_{jl}(Y_l, \mathbf{t}_0)$$

under the conjugate density $h = \prod_l h_l$ and $u_0 = u(\mathbf{t}_0)$.

Chapter 3

Tests on the value of regression quantile

Let us assume we have the observations $Y_1, \ldots, Y_n$ following the model (2.1). Suppose we want to test hypothesis

$$H_0 : \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha 0} \in \mathbb{R}^{p+1} \text{ against } H_1 : \text{ non } H_0. \tag{3.1}$$

In case the distribution of the errors $e_1, \ldots, e_n$ is known, classical parametric tests based on likelihood, like Wald test, Rao score test or likelihood ratio test, can be used. Although these tests possess asymptotic optimal properties under the assumed model, they are not only nonrobust to small departures from assumed distribution, but their relative error is of order $n^{-1/2}$. One of the robust alternatives to these tests are M-tests or eventually M-test of the Wald type. Unfortunately, these tests require estimation of variance of the M-estimator.

Rank test for regression quantiles proposed by Gutenbrunner et al. (1993) are also available, Koenker (1994) later inverted this test in order to compute confidence intervals for $\boldsymbol{\beta}_\alpha$.

Another approach to this problem might be saddlepoint tests. As saddlepoint approximations for the density of an estimator have many desirable properties, later have been developed also saddlepoint tests for M-estimators with relative error of order n^{-1}. We will deal with sample quantiles and regression quantiles that are M-estimators with ρ-function having simple form that allows us to write the test statistic by an explicit formula, not only in parametric case but also in nonparametric case.

In this chapter we will propose two new tests for hypothesis H_0, the first one based on the asymptotic distribution of averaged regression quantiles that will require the estimation of quantile density function and the second one will be derived using theory of saddlepoint tests for M-estimators. Combination of simplicity of ψ-function for regression quantile as an M-estimator and asymptotic properties of saddlepoint tests allow us to construct a test statistic with an explicit formula that makes proposed test much easier to use and works very well even under small sample sizes. The performance of the proposed tests will be numerically illustrated and compared with other tests. This chapter is part of the manuscript (Ronchetti and Sabolová, 2014) which is about to be submitted.

3.1 Test based on averaged regression quantiles

The scalar statistic

$$\bar{\mathbf{x}}_n^T \hat{\boldsymbol{\beta}}_\alpha,$$

where $\bar{\mathbf{x}}_n = \sum_{i=1}^n x_i$ is called the averaged regression quantile. The proposed test statistic is based on the asymptotic distribution of averaged regression quantiles. Even though it requires the estimation of the quantile density function, unlike in the case of classical tests based on likelihood, it does not require estimation of the variance matrix.

Theorem 6. *Consider the regression model* $Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + e_i$, $i = 1, \ldots, n$, *where* e_i *are i.i.d. with distribution function* F. *Suppose that the distribution function* F *is continuous and twice differentiable in neighborhood of* $F^{-1}(\alpha)$ *and that* $F'(F^{-1}(\alpha)) = f(F^{-1}(\alpha)) > 0$, $0 < \alpha < 1$. *Let the following regularity conditions on the matrix* $\mathbf{X}$ *be satisfied*

(B1) $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}^T\mathbf{X} = \mathbf{Q}$, *where* $\mathbf{Q}$ *is a positive definite matrix*

(B2) $\frac{1}{n}\sum_{i=1}^n x_{ij} = O(1)$ *as* $n \to \infty$ *for* $j = 1, \ldots, p$.

Consider the null hypothesis

$$H_0 : \boldsymbol{\beta}_\alpha = \boldsymbol{\beta}_{\alpha_0} \in \mathbb{R}^p$$

against alternative

$$H_1 : \text{ non } H_0.$$

The test statistic

$$\sqrt{n}\frac{f(F^{-1}(\alpha))}{\sqrt{\alpha(1-\alpha)}}\bar{x}^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha_0})$$

is under H_0 *asymptotically* $N(0,1)$ *distributed.*

Proof. Gutenbrunner and Jurečková (1992) showed that under null hypothesis the regression quantile admits the following second-order representation

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha\right) = \frac{1}{\sqrt{n}f(F^{-1}(\alpha))}Q_n^{-1}\sum_{i=1}^n \mathbf{x}_i\psi_\alpha(E_{i\alpha}) + O_P(n^{-1/4}),$$

where $E_{i\alpha} = e_i - F^{-1}(\alpha)$ and

$$\psi_\alpha(x) = \begin{cases} \alpha & : x > 0 \\ \alpha - 1 & : x \le 0 \end{cases}$$

and from this it follows that

$$\sqrt{n}\bar{\mathbf{x}}^T\left(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha\right) = \frac{1}{\sqrt{n}}f(F^{-1}(\alpha))\sum_{i=1}^n \left(\alpha - I[e_i < F^{-1}(\alpha)]\right) + O_P(n^{-1/4}).$$

From the properties of sample quantiles we get that $I[e_i < F^{-1}(\alpha)] \sim Alt(\alpha)$, thus

$$\begin{aligned} \mathsf{E}\left(\alpha - I[e_i < F^{-1}(\alpha)]\right) =& 0 \\ \operatorname{var}\left(\alpha - I[e_i < F^{-1}(\alpha)]\right) =& \alpha(1-\alpha). \end{aligned}$$

and by using Taylor moment expansion and central limit theorem we get that under null hypothesis

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\alpha - I[e_i < F^{-1}(\alpha)]\right) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \alpha(1-\alpha)).$$

□

Remark 6. *Computation of test statistic requires estimation of quantile density function that might be done using method already described in the previous chapter. Test based on averaged regression quantiles simplifies the computation of the test statistic and speeds up the numerical simulations as it is not necessary to work with matrices.*

The performance and robustness of the test in comparison with other known tests will be studied later in this chapter.

3.2 Saddlepoint tests for sample and regression quantiles

As sample and regression quantiles can be rewritten as M-estimators, it is possible to derive the test statistic for hypothesis (3.1) based on the theory of the saddlepoint tests outlined in the first chapter. Even though the simple form of the ρ-function and its directional derivatives yield explicit formula for the test statistics, these functions do not satisfy conditions assumed in the theorems in Ronchetti et al. (2003) and Ma and Ronchetti (2011) and thus it is not possible to base the asymptotic distribution of the proposed statistics on the available theory. Therefore in order to derive the test statistics, we will follow the development introduced in aforementioned papers, but the proofs of the asymptotic distribution under null hypothesis will have to be rewritten using other techniques.

We will first derive the test statistic for the sample quantile $F^{-1}(\alpha)$ and later concentrate on the test on the value of regression quantile. We are particularly interested in the parametric and nonparametric tests for small sample sizes. We will follow the development introduced in Ronchetti et al. (2003) and Ma and Ronchetti (2011). Test statistic is based on the exponent in the saddlepoint approximation to the density of the function of M-estimators. Although this statistic is analogous to the log likelihood ratio in the parametric case, the relative error is only of order n^{-1}. In case the distribution of the observations is unknown, we will use empirical likelihood statistic.

3.2.1 Test on the value of $F^{-1}(\alpha)$ in location model

Let us first consider the simplest case of regression quantiles, i.e. sample quantiles. The quantile function is often intractable (i.e. for normally distributed random variable), therefore it is useful to construct the test on the value of α-quantile.

Let $Y_1, \ldots, Y_n$ be independent identically distributed sample of random variables with distribution function F and underlying density f. Denote for $0 < \alpha < 1$

$$\beta_\alpha(F) = F^{-1}(\alpha)$$

an α-quantile of distribution F, i.e. M-estimator with ψ-function equal to

$$\psi(y, \beta_\alpha) = \alpha I[y > \beta_\alpha] - (1-\alpha) I[y < \beta_\alpha] = \alpha - I[y < \beta_\alpha]. \tag{3.2}$$

Denote by $\hat{\beta}_\alpha$ its empirical counterpart, i.e.

$$\hat{\beta}_\alpha = F_n^{-1}(\alpha),$$

where

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}[Y_i \leq y]$$

is the empirical distribution function. Suppose we want to test the null hypothesis

$$H_0 : \beta_\alpha = \beta_{\alpha 0}$$

against alternative

$$H_1 : \beta_\alpha \neq \beta_{\alpha 0}.$$

We will derive test statistic for parametric case when $F(.)$ is known and also for nonparametric case when it is necessary to construct hypothetical distribution function $\hat{F}_0$.

Parametric case

Theorem 7. *Let $Y_1, \ldots, Y_n$ be i.i.d. with the distribution F_0, with corresponding density f_0 and let $f_0(F^{-1}(\alpha)) > 0$ for $\alpha \in (0,1)$. Consider null hypothesis*

$$H_0 : \beta_\alpha = \beta_{\alpha 0}$$

against two-sided alternative

$$H_1 : \beta_\alpha \neq \beta_{\alpha 0}.$$

The saddlepoint test statistic has the form

$$2n \log \left(\left(\frac{F_0(\hat{\beta}_\alpha)}{\alpha} \right)^{-\alpha} \left(\frac{1 - F_0(\hat{\beta}_\alpha)}{1-\alpha} \right)^{-(1-\alpha)} \right)$$

and under null distribution is asymptotically χ^2_1-distributed.

Proof. Let F_0 be the distribution of observations $Y_1, \ldots, Y_n$ under null hypothesis. Let us compute cumulant generating function of score function (3.2) corresponding to α-quantile under H_0 :

$$\begin{aligned} K_\psi(\lambda, \beta_\alpha) &= \log \mathsf{E} e^{\lambda\psi(Y,\beta_\alpha)} \\ &= \log e^{\lambda\alpha} \mathsf{E} e^{-\lambda I\{y<\beta_\alpha\}} \\ &= \lambda\alpha + \log \underbrace{\int_{-\infty}^{\infty} e^{-\lambda I\{y<\beta_\alpha\}} f_0(y) dy}_{\int_{-\infty}^{\beta} e^{-\lambda} f_0(y)dy + \int_{\beta}^{\infty} f_0(y)dy} \\ &= \lambda\alpha + \log(e^{-\lambda} F_0(\beta_\alpha) + 1 - F_0(\beta_\alpha)). \end{aligned}$$

To obtain saddlepoint $\lambda(\beta_\alpha)$ of K_ψ, we have to solve the equation

$$\frac{\partial K_\psi(\lambda, \beta_\alpha)}{\partial \lambda} = \alpha - \frac{e^{-\lambda} F_0(\beta_\alpha)}{e^{-\lambda} F_0(\beta_\alpha) + 1 - F_0(\beta_\alpha)} = 0$$

with respect to λ, the solution being

$$\lambda(\beta_\alpha) = \log \frac{F_0(\beta_\alpha)(1-\alpha)}{\alpha(1-F_0(\beta_\alpha))}$$

and therefore by (1.6) and (1.7) we can rewrite test statistic in a following way

$$\begin{aligned} h(\hat{\beta}_\alpha) &= \sup_\lambda \{-K_\psi(\lambda, \hat{\beta}_\alpha)\} \\ &= -K_\psi(\lambda(\hat{\beta}_\alpha), \hat{\beta}_\alpha) \\ &= -\alpha \log \frac{\alpha(1-F_0(\hat{\beta}_\alpha)}{F(\hat{\beta}_\alpha)(1-\alpha)} + \log \frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha} \\ &= \log \left(\left(\frac{F_0(\hat{\beta}_\alpha)}{\alpha} \right)^{-\alpha} \left(\frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha} \right)^{-(1-\alpha)} \right) \end{aligned}$$

Let us now approximate the asymptotic distribution of the derived test statistic $2nh(\hat{\beta}_\alpha)$. We will use only the first two terms Taylor expansion for function F_0

$$F_0(\hat{\beta}_\alpha) = F_0(\beta_\alpha) + (\hat{\beta}_\alpha - \beta_\alpha) F_0'(\beta_\alpha) = \alpha + (\hat{\beta}_\alpha - \beta_\alpha) f_0(\beta_\alpha)$$

and the test statistic can be rewritten in a following way (using Taylor expansion of function $\log(1+x)$ for $|x| < 1$)

$$\begin{aligned} 2nh(\hat{\beta}_\alpha) =& 2n \log \left(\left(\frac{F_0(\hat{\beta}_\alpha)}{\alpha} \right)^{-\alpha} \left(\frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha} \right)^{-(1-\alpha)} \right) \\ =& -2n \left(\alpha \log \frac{F_0(\hat{\beta}_\alpha)}{\alpha} + (1-\alpha) \log \frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha} \right) \\ \approx& -2n \left(\alpha \log \left(1 + \frac{(\hat{\beta}_\alpha - \beta_\alpha) f_0(\beta_\alpha)}{\alpha} \right) + (1-\alpha) \log \left(1 + \frac{(\hat{\beta}_\alpha - \beta_\alpha) f_0(\beta_\alpha)}{1-\alpha} \right) \right) \end{aligned}$$

$$
\begin{aligned}
\approx & -2n\left((\hat{\beta}_\alpha-\beta_\alpha)f_0(\beta_\alpha)-\frac{1}{2}\frac{(\hat{\beta}_\alpha-\beta_\alpha)^2f_0^2(\beta_\alpha)}{\alpha}+(\beta_\alpha-\hat{\beta}_\alpha)f_0(\beta_\alpha)\right.\\
& \left.-\frac{1}{2}\frac{(\hat{\beta}_\alpha-\beta_\alpha)^2f_0^2(\beta_\alpha)}{1-\alpha}\right)\\
= & n\frac{(\hat{\beta}_\alpha-\beta_\alpha)^2f_0^2(\beta_\alpha)}{\alpha(1-\alpha)}.
\end{aligned}
$$

The sample quantile is asymptotically normally distributed

$$
\sqrt{n}(\hat{\beta}_\alpha-\beta_\alpha)\xrightarrow[n\to\infty]{\mathcal{D}}N\left(0,\frac{\alpha(1-\alpha)}{f_0^2(\beta_\alpha)}\right),
$$

thus

$$
n\frac{(\hat{\beta}_\alpha-\beta_\alpha)^2f_0^2(\beta_\alpha)}{\alpha(1-\alpha)}\xrightarrow[x\to\infty]{\mathcal{D}}\chi_1^2.
$$

Hence under null hypothesis

$$
2nh(\hat{\beta}_\alpha)=2n\log\left(\left(\frac{F_0(\hat{\beta}_\alpha)}{\alpha}\right)^{-\alpha}\left(\frac{1-F_0(\hat{\beta}_\alpha)}{1-\alpha}\right)^{-(1-\alpha)}\right)
$$

is asymptotically χ_1^2-distributed. □

Nonparametric case

Let us denote

$$
\begin{aligned}
F_n(\beta)= & \frac{1}{n}\sum_{i=1}^{n}I[y_i-\beta<0]\\
I_i(\beta)= & \mathrm{I}[y_i-\beta<0].
\end{aligned}
$$

Theorem 8. *The saddlepoint test statistic has the form*

$$
-2n\left\{\alpha\log\frac{\sum_{i=1}^n w_iI_i}{\alpha}+(1-\alpha)\log\frac{\sum_{i=1}^n w_i(1-I_i)}{1-\alpha}\right\},
$$

where

$$
w_i=\frac{\left\{\frac{\alpha(1-F_n(\beta_{\alpha0}))}{F_n(\beta_{\alpha0})(1-\alpha)}\right\}^{I\{y_i<\beta_{\alpha0}\}}}{\sum_{j=1}^n\left\{\frac{\alpha(1-F_n(\beta_{\alpha0}))}{F_n(\beta_{\alpha0})(1-\alpha)}\right\}^{I\{y_j<\beta_{\alpha0}\}}},\qquad i=1,\ldots,n
$$

and is under null hypothesis χ_1^2-distributed.

Let us now state and prove the lemma necessary for proving the asymptotic distribution of the test statistic.

Lemma 2. *Under null hypothesis $H_0 : \beta_\alpha = \beta_{\alpha 0}$ it holds*

$$\sup_{i=1,\dots,n} \left| w_i - \frac{1}{n} \right| \xrightarrow[n\to\infty]{0} \quad a.\ s.$$

Proof. Under null hypothesis we have

$$\begin{aligned}
\left| w_i - \frac{1}{n} \right| &= \left| \frac{\left\{ \frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)} \right\}^{I\{y_i<\beta_{\alpha 0}\}}}{\sum_{j=1}^n \left\{ \frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)} \right\}^{I\{y_j<\beta_{\alpha 0}\}}} - \frac{1}{n} \right| \\
&= \frac{1}{n} \frac{1}{\frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)} \sum_{j=1}^n I[y_j < \beta_{\alpha 0}] + n - \sum_{j=1}^n I[y_j < \beta_{\alpha 0}]} \\
&\times \left| n \frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)} I[y_i < \beta_{\alpha 0}] + n(1 - I[y_i < \beta_{\alpha 0}]) \right. \\
&\left. - \sum_{j=1}^n \frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)} I[y_j < \beta_{\alpha 0}] + \sum_{j=1}^n (1 - I[y_j < \beta_{\alpha 0}]) \right| \\
&= \frac{1}{n} \left| \frac{n(I[y_i < \beta_{\alpha 0}] - F_n(\beta_{\alpha 0})) \frac{\alpha - F_n(\beta_{\alpha 0})}{(1-\alpha)F_n(\beta_{\alpha 0})}}{n\left(1 + \frac{\alpha - F_n(\beta_{\alpha 0})}{1-\alpha}\right)} \right|
\end{aligned}$$

and by Glivenko-Cantelli theorem under null hypothesis $\sup |F_n(\beta_{\alpha 0}) - \alpha| \to 0$ almost surely. This concludes the proof. □

Let us now proceed to the proof of the Theorem 8.

Proof. As the distribution of observations is unknown, it is necessary to find the empirical hypothetical distribution function $\hat{F}_0 = (w_1, \dots, w_n)^T$ using (1.8). The weights w_i for sample quantile are of the form

$$w_i = \frac{e^{\mu(\beta_{\alpha 0})\psi(y_i,\beta_{\alpha 0})}}{\sum_{j=1}^n e^{\mu(\beta_{\alpha 0})\psi(y_j,\beta_{\alpha 0})}}, \tag{3.3}$$

where $\mu(\beta_{\alpha 0})$ is the solution of the following maximization problem

$$\begin{aligned}
\mu(\beta_{\alpha 0}) &= \operatorname{argmax} \left\{ -\log \left(\frac{1}{n} \sum_{i=1}^n e^{\mu\psi(y_i,\beta_{\alpha 0})} \right) \right\} \\
&= \operatorname{argmax} \left\{ -\log \left(\frac{1}{n} \sum_{i=1}^n e^{\mu}\, \mathrm{I}[y_i \le \beta_{\alpha 0}] + \frac{1}{n} \sum_{i=1}^n (1 - \mathrm{I}[y_i \le \beta_{\alpha 0}]) \right) \right\} \\
&= \operatorname{argmax}\{-\mu\alpha - \log(e^{-\mu} F_n(\beta_{\alpha 0}) + 1 - F_n(\beta_{\alpha 0}))\}.
\end{aligned} \tag{3.4}$$

The solution $\mu(\beta_\alpha)$ equals

$$\mu(\beta_\alpha) = \log \frac{F_n(\beta_\alpha)(1-\alpha)}{\alpha(1 - F_n(\beta_\alpha))}.$$

By inserting $\mu(\beta_{\alpha 0})$ into (1.8) we get the formula for components of $\hat{F}_0$

$$w_i = \frac{\left\{\frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)}\right\}^{I\{y_i<\beta_{\alpha 0}\}}}{\sum_{j=1}^n \left\{\frac{\alpha(1-F_n(\beta_{\alpha 0}))}{F_n(\beta_{\alpha 0})(1-\alpha)}\right\}^{I\{y_j<\beta_{\alpha 0}\}}}, \qquad i = 1, \ldots, n.$$

Then the test statistic has the form

$$h(\hat{\beta}_\alpha) = \sup_\lambda \left\{ -\log \sum_{i=1}^n w_i e^{\lambda\psi(y_i,\hat{\beta}_\alpha)} \right\}$$

and can be rewritten using (1.8). The saddlepoint $\lambda(\hat{\beta}_\alpha)$ solves the equation

$$\frac{\partial}{\partial\lambda}\left\{ -\log \sum_{i=1}^n w_i e^{\lambda\psi(y_i,\hat{\beta}_\alpha)} \right\} = -\frac{\sum_{i=1}^n w_i e^{(\alpha - I\{y_i<\hat{\beta}_\alpha\})\lambda}(\alpha - I[y_i < \hat{\beta}_\alpha])}{\sum_{i=1}^n w_i e^{(\alpha - I\{y_i<\hat{\beta}_\alpha\})\lambda}} = 0$$

and is equal to

$$\lambda(\hat{\beta}_\alpha) = \log \frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^n w_i \,\mathrm{I}[y_i \le \hat{\beta}_\alpha]}{\sum_{i=1}^n w_i (1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha])}.$$

This leads to the following formula for the test statistic

$$\begin{aligned}
h(\hat{\beta}_\alpha) &= -\log \sum_{i=1}^n w_i e^{\lambda\psi(y_i,\hat{\beta}_\alpha)} \\
&= -\log \left(\sum_{i=1}^n w_i \,\mathrm{I}[y_i \le \hat{\beta}_\alpha] e^{(\alpha-1)\log\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_i \mathrm{I}[y_i\le\hat{\beta}_\alpha]}{\sum_{i=1}^n w_i(1-\mathrm{I}[y_i\le\hat{\beta}_\alpha])}} + \right. \\
&\left. \sum_{i=1}^n w_i(1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha]) e^{\alpha\log\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_i \mathrm{I}[y_i\le\hat{\beta}_\alpha]}{\sum_{i=1}^n w_i(1-\mathrm{I}[y_i\le\hat{\beta}_\alpha])}} \right) \\
&= -\log \sum_{i=1}^n \frac{1}{1-\alpha} w_i (1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha]) \left(\frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^n w_i \,\mathrm{I}[y_i \le \hat{\beta}_\alpha]}{\sum_{i=1}^n w_i (1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha])} \right)^\alpha \\
&= -\log \left\{ \left(\frac{\sum_{i=1}^n w_i (1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha])}{1-\alpha} \right)^{1-\alpha} \left(\frac{\sum_{i=1}^n w_i \,\mathrm{I}[y_i \le \hat{\beta}_\alpha]}{\alpha} \right)^\alpha \right\}.
\end{aligned}$$

Then the test statistic has the form

$$\begin{aligned}
-2n\log &\left\{ \left(\frac{\sum_{i=1}^n w_i (1 - \mathrm{I}[y_i \le \hat{\beta}_\alpha])}{1-\alpha} \right)^{1-\alpha} \left(\frac{\sum_{i=1}^n w_i \,\mathrm{I}[y_i \le \hat{\beta}_\alpha]}{\alpha} \right)^\alpha \right\} \\
&= -2n \left\{ \alpha \log \frac{\sum_{i=1}^n w_i I_i}{\alpha} + (1-\alpha)\log \frac{\sum_{i=1}^n w_i (1 - I_i)}{1-\alpha} \right\},
\end{aligned}$$

where logarithmic function can be approximated by the first two terms of the Taylor expansion

$$\begin{aligned}\log\left(\frac{\sum_{i=1}^n w_iI_i}{\alpha}\pm 1\right) =& \log\left(1+\frac{\sum_{i=1}^n w_iI_i-\alpha}{\alpha}\right)\\ =&\frac{\sum_{i=1}^n w_iI_i-\alpha}{\alpha}-\frac{1}{2}\frac{(\sum_{i=1}^n w_iI_i-\alpha)^2}{\alpha^2}+...\\ \log\left(\frac{\sum_{i=1}^n w_i(1-I_i)}{1-\alpha}\pm 1\right) =& \log\left(1+\frac{\sum_{i=1}^n w_iI_i-\alpha}{\alpha}\right)\\ =&\frac{\sum_{i=1}^n w_iI_i-\alpha}{\alpha}-\frac{1}{2}\frac{(\sum_{i=1}^n w_iI_i-\alpha)^2}{\alpha^2}+....\end{aligned}$$

Then the test statistic can be approximated by

$$n\frac{(\sum_{i=1}^n w_iI_i-\alpha)^2}{\alpha(1-\alpha)}.$$

Under H_0 we have $I_i \sim Alt(\alpha)$. Then

$$\mathsf{E}(\sum_{i=1}^n w_iI_i-\alpha)\approx 0$$
$$\mathrm{var}(\sum_{i=1}^n w_iI_i-\alpha)\approx\frac{\alpha(1-\alpha)}{n}$$

and therefore (using CLT) under null hypothesis

$$\begin{aligned}2nh(\hat\beta_\alpha) =& 2n\log\left\{\left(\frac{\sum_{i=1}^n w_i(1-\mathrm{I}[y_i\le\hat\beta_\alpha])}{1-\alpha}\right)^{-(1-\alpha)}\left(\frac{\sum_{i=1}^n w_i\,\mathrm{I}[y_i\le\hat\beta_\alpha]}{\alpha}\right)^{-\alpha}\right\}\\ &\xrightarrow[n\to\infty]{\mathcal{D}}\chi_1^2.\end{aligned}$$

□

3.2.2 Test on the value of regression quantile based on saddlepoint techniques

Let $Y_1,\dots,Y_n$ be observations following the model

$$Y_i=\mathbf{X}_i^T\boldsymbol{\beta}+e_i,\qquad i=1,\dots,n$$

where $e_i\sim\frac{1}{\sigma}g(.)$ and $(Y_i,\mathbf{X}_i)$ are i.i.d. with density $\frac{1}{\sigma}g\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x}_i)$ with $k(\mathbf{x}_i)$ being the density of $\mathbf{X}_i$. We will first consider more general situation with random matrix $\mathbf{X}$ and later simplify the result for fixed matrix $\mathbf{X}$. We want to test simple hypothesis

$$H_0:\boldsymbol{\beta}_\alpha=\boldsymbol{\beta}_{\alpha_0}$$

against two sided alternative

$$H_1:\boldsymbol{\beta}_\alpha\neq\boldsymbol{\beta}_{\alpha_0}.$$

Parametric case

Theorem 9. *Consider the model (2.1), where* $e_i \sim \frac{1}{\sigma}g(.)$ *and* $(Y_i, \mathbf{X}_i)$ *are i.i.d. with density* $\frac{1}{\sigma}g\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x}_i)$ *with* $k(\mathbf{x}_i)$ *being the density of* $\mathbf{X}_i$. *Assume conditions (A1) and (A2) in Theorem 2 are satisfied. The saddlepoint test statistic has the form*

$$2nh(\hat{\boldsymbol{\beta}}_\alpha) = -2n\log \mathsf{E}_{\mathbf{x}}\left[\left(\frac{G\left(\mathbf{x}^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})\right)}{\alpha}\right)^\alpha\left(\frac{1-G\left(\mathbf{x}^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})\right)}{1-\alpha}\right)^{1-\alpha}\right].$$

For fixed matrix $\mathbf{X}$ *the test statistic equals*

$$-2\sum_{i=1}^n \alpha\log\left(1+\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})f_0(\boldsymbol{\beta}_{\alpha 0})}{\alpha}\right)+(1-\alpha)\log\left(1+\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})f_0(\boldsymbol{\beta}_{\alpha 0})}{1-\alpha}\right)$$

and under H_0 *is asymptotically* χ^2_{p+1}*-distributed.*

Proof. Cumulant generating function of score function corresponding to regression quantile can be expressed in a following way

$$\begin{aligned}
K_\psi(\boldsymbol{\lambda},\boldsymbol{\beta}_\alpha) &= \log \mathsf{E}e^{\boldsymbol{\lambda}\psi(Y_i,\boldsymbol{\beta}_\alpha)\mathbf{x}_i}\\
&= \log \mathsf{E}e^{\boldsymbol{\lambda}^T\mathbf{x}_i\left(\alpha-\mathrm{I}\left[\frac{Y_i-\mathbf{x}_i^T\boldsymbol{\beta}_\alpha}{\sigma}<0\right]\right)}\\
&= \log\int\int\frac{1}{\sigma}e^{\alpha\boldsymbol{\lambda}^T\mathbf{x}_i}e^{-\boldsymbol{\lambda}^T\mathbf{x}_i\mathrm{I}\left[\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}_\alpha}{\sigma}<0\right]}g\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x}_i)\,dy_i\,d\mathbf{x}_i\\
&= \log\int\left[\int_{-\infty}^{\mathbf{x}_i^T\boldsymbol{\beta}_\alpha}\frac{1}{\sigma}e^{\alpha\boldsymbol{\lambda}^T\mathbf{x}_i-\boldsymbol{\lambda}^T\mathbf{x}_i}g\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x}_i)\,dy_i+\right.\\
&\quad\left.\int_{\mathbf{x}_i^T\boldsymbol{\beta}_\alpha}^{\infty}\frac{1}{\sigma}e^{\alpha\boldsymbol{\lambda}^T\mathbf{x}_i}g\left(\frac{y_i-\mathbf{x}_i^T\boldsymbol{\beta}}{\sigma}\right)k(\mathbf{x}_i)\,dy_i\right]d\mathbf{x}_i\\
&= \log\int\left\{e^{\alpha\boldsymbol{\lambda}^T\mathbf{x}_i}k(\mathbf{x}_i)\left(e^{-\boldsymbol{\lambda}^T\mathbf{x}_i}G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha-\boldsymbol{\beta})}{\sigma}\right)\right.\right.\\
&\quad\left.\left.+1-G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha-\boldsymbol{\beta})}{\sigma}\right)\right)\right\}d\mathbf{x}_i
\end{aligned}$$

In order to compute saddlepoint the derivative of $K_\psi(\boldsymbol{\lambda},\boldsymbol{\beta}_\alpha)$ with respect to $\boldsymbol{\lambda}$ has to be computed

$$\begin{aligned}
\frac{\partial K_\psi(\boldsymbol{\lambda},\boldsymbol{\beta}_\alpha)}{\partial\boldsymbol{\lambda}} =& \frac{\partial}{\partial\boldsymbol{\lambda}}\log\int\left\{e^{\alpha\boldsymbol{\lambda}^T\mathbf{x}_i}k(\mathbf{x}_i)\left(e^{-\boldsymbol{\lambda}^T\mathbf{x}_i}G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha-\boldsymbol{\beta})}{\sigma}\right)\right.\right.\\
&\left.\left.+1-G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha-\boldsymbol{\beta})}{\sigma}\right)\right)\right\}d\mathbf{x}_i
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\int \left\{ \frac{\partial}{\partial \boldsymbol{\lambda}} e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) + 1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i}{\int \left\{ e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma}) + 1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i} \\
&= \left\{ \int \left\{ \alpha \mathbf{x}_i e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right. \right. \right. \\
&\quad \left. +1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) + \\
&\quad \left. \left. e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(-\mathbf{x}_i e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i \right\} \times \\
&\quad \left\{ \int \left\{ e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right. \right. \right. \\
&\quad \left. \left. \left. +1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_\alpha - \boldsymbol{\beta})}{\sigma} \right) \right) \right\} d\mathbf{x}_i \right\}^{-1}
\end{aligned}
$$

By solving

$$\frac{\partial K_\psi(\boldsymbol{\lambda}, \boldsymbol{\beta}_\alpha)}{\partial \boldsymbol{\lambda}} = 0$$

we get the function of saddlepoint $\boldsymbol{\lambda}^T \mathbf{x}_i$

$$\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha)^T \mathbf{x}_i = -\log \left\{ \frac{\alpha}{1-\alpha} \frac{1 - G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)} \right\}, \qquad i = 1, \ldots, n.$$

Therefore

$$-K_\psi(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha), \hat{\boldsymbol{\beta}}_\alpha) \;=\; -\log \int \left(\frac{G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{\alpha} \right)^\alpha \left(\frac{1 - G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{1-\alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i$$

and hence the test statistic has the form

$$2nh(\hat{\boldsymbol{\beta}}_\alpha) = -2n \log \int \left(\frac{G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{\alpha} \right)^\alpha \left(\frac{1 - G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{1-\alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i.$$

In case matrix $\mathbf{X}$ is known, the test statistic can be rewritten as a mean over rows of matrix X

$$
\begin{aligned}
&-2 \sum_{i=1}^n \log \int \left(\frac{G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{\alpha} \right)^\alpha \left(\frac{1 - G\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta})}{\sigma} \right)}{1-\alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i \\
&= -2 \sum_{i=1}^n \alpha \log \left(1 + \frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) f_0(\boldsymbol{\beta}_{\alpha 0})}{\alpha} \right) + (1-\alpha) \log \left(1 + \frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha 0}) f_0(\boldsymbol{\beta}_{\alpha 0})}{1-\alpha} \right)
\end{aligned}
$$

$$
\begin{aligned}
\approx & -2\sum_{i=1}^{n}\alpha\left(\frac{\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g_0(\boldsymbol{\beta}_{\alpha 0})}{\alpha}-\frac{1}{2}\frac{(\mathbf{x}_i^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta}_{\alpha 0})g_0(\boldsymbol{\beta}_{\alpha 0}))^2}{\alpha^2}\right)\\
& +(1-\alpha)\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha)g_0(\boldsymbol{\beta}_{\alpha 0})}{\alpha}-\frac{1}{2}\frac{(\mathbf{x}_i^T(\boldsymbol{\beta}_{\alpha 0}-\hat{\boldsymbol{\beta}}_\alpha)g_0(\boldsymbol{\beta}_{\alpha 0}))^2}{(1-\alpha)^2}\right)\\
= & -2\sum_{i=1}^{n}\frac{(\mathbf{x}_i^T(\boldsymbol{\beta}_{\alpha_0}-\hat{\boldsymbol{\beta}}_\alpha)g_0(\boldsymbol{\beta}_{\alpha_0}))^2}{\alpha(1-\alpha)}\\
= & \frac{1}{n}\frac{g_0^2(\boldsymbol{\beta}_{\alpha_0})}{\alpha(1-\alpha)}\sum_{i=1}^{n}\left(\sqrt{n}\mathbf{x}_i^T(\boldsymbol{\beta}_{\alpha_0}-\hat{\boldsymbol{\beta}}_\alpha)\right)^2.
\end{aligned}
$$

The proof can be completed using the asymptotic distribution of the regression quantile (see Theorem 2)

$$
\sqrt{n}(\boldsymbol{\beta}_{\alpha_0}-\hat{\boldsymbol{\beta}}_\alpha)\xrightarrow[n\to\infty]{\mathcal{D}}N_{p+1}\left(0,\frac{\alpha(1-\alpha)}{g_0(\boldsymbol{\beta}_{\alpha_0})^2}\mathbf{D}_0^{-1}\right),
$$

thus the asymptotic distribution of a test statistic is χ^2_{p+1}. □

Nonparametric case

In nonparametric setup we will again assume e_i are i.i.d. with distribution function F and in the following computations work with residuals $r_i = y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_\alpha$, $i = 1, \dots, n$. Let us denote

$$
\begin{aligned}
I_i &= \mathrm{I}[r_i<0]\\
I_{ij}(\boldsymbol{\beta}) &= \mathrm{I}[r_i+\mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha-\boldsymbol{\beta})<0]\\
F_n^j(\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n}I_{ij}(\boldsymbol{\beta})\\
w_{ij} &= \frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}}{\sum_{k=1}^{n}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{kj}(\boldsymbol{\beta}_{\alpha_0})}}.
\end{aligned}
$$

Theorem 10. *Let $Y_1, \dots, Y_n$ be observations following the model*

$$
Y_i = \mathbf{X}_i^T\boldsymbol{\beta} + e_i, \qquad i = 1, \dots, n
$$

where e_i are i.i.d. with distribution function F, $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^T \in R^{p+1}$, $\boldsymbol{\beta} \in R^{p+1}$. The saddlepoint test statistic equals

$$
-2\sum_{j=1}^{n}\alpha\log\frac{\sum_{i=1}^{n}w_{ij}I_i}{\alpha}+(1-\alpha)\log\frac{\sum_{i=1}^{n}w_{ij}(1-I_i)}{1-\alpha}.
$$

Proof. We will follow the development introduced in Ronchetti and Welsh (1994) and consider empirical distribution functions of residuals

$$r_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\alpha, \qquad i = 1, \ldots, n$$

and functions

$$\begin{aligned} I_{ij}(\boldsymbol{\beta}) =& \mathrm{I}[r_i + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}) < 0] \\ F_n^j(\boldsymbol{\beta}) =& \frac{1}{n}\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}). \end{aligned}$$

Because it is not possible to derive explicit formula for $\boldsymbol{\mu}$ (see (1.11)), we will use formula for $\mathbf{x}_j^T\boldsymbol{\mu}$

$$\mathbf{x}_j^T\boldsymbol{\mu} = \log \frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha_0})}{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}$$

that solves

$$\sum_{i=1}^n\sum_{j=1}^n (\alpha - \mathrm{I}[r_i + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha_0})])\mathbf{x}_j e^{(\alpha - \mathrm{I}[r_i + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha_0})])\mathbf{x}_j^T\boldsymbol{\mu}} = 0,$$

because

$$\sum_{i=1}^n\sum_{j=1}^n (\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha_0})}{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{\alpha - I_{ij}(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j = 0$$

$$\sum_{i=1}^n\sum_{j=1}^n \alpha\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha_0})}{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{-I_{ij}(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j =$$

$$= \sum_{i=1}^n\sum_{j=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha_0})\left(\frac{1-\alpha}{\alpha}\frac{F_n^j(\boldsymbol{\beta}_{\alpha_0})}{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{-I_{ij}(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j$$

$$\sum_{i=1}^n\sum_{j=1}^n \alpha\frac{\alpha}{1-\alpha}\frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}I_{ij}(\boldsymbol{\beta}_{\alpha_0})\mathbf{x}_j + \sum_{i=1}^n\sum_{j=1}^n \alpha(1 - I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j =$$

$$= \sum_{i=1}^n\sum_{j=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha_0})\frac{\alpha}{1-\alpha}\frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j$$

$$\sum_{i=1}^n\sum_{j=1}^n \alpha\mathbf{x}_j(1 - I_{ij}(\boldsymbol{\beta}_{\alpha_0})) = \sum_{i=1}^n\sum_{j=1}^n (1-\alpha)I_{ij}(\boldsymbol{\beta}_{\alpha_0})\frac{\alpha}{1-\alpha}\frac{1 - F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j$$

$$n\sum_{j=1}^n \mathbf{x}_j - \sum_{j=1}^n\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha_0})\mathbf{x}_j = \sum_{j=1}^n\sum_{i=1}^n \frac{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j - \sum_{j=1}^n\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha_0})\mathbf{x}_j$$

$$n\sum_{j=1}^n \mathbf{x}_j = \sum_{j=1}^n \frac{1}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\mathbf{x}_j \underbrace{\sum_{i=1}^n I_{ij}(\boldsymbol{\beta}_{\alpha_0})}_{nF_n^j(\boldsymbol{\beta}_{\alpha_0})}$$

Then we can write

$$\begin{aligned}
K_\psi(\boldsymbol{\lambda},\boldsymbol{\beta}) &= \frac{1}{n}K^j(\boldsymbol{\lambda},\boldsymbol{\beta}) \\
&= \frac{1}{n}\sum_{j=1}^n \left\{\log\left[e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^T\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}}\right]-\log\sum_{i=1}^n e^{(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}}\right\} \\
\frac{\partial K^j(\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\boldsymbol{\lambda}} &= \frac{\partial}{\partial\boldsymbol{\lambda}}\left\{\log\sum_{i=1}^n e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^T\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}}\right\} \\
&= \frac{\sum_{i=1}^n(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^T\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}}}{\sum_{i=1}^n e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^T\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}}}
\end{aligned}$$

Therefore the test statistic has the form

$$2n\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) = 2n\sup_{\boldsymbol{\lambda}} K_\psi(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha),\hat{\boldsymbol{\beta}}_\alpha),$$

where $\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha)$ satisfies

$$\begin{aligned}
\frac{\partial K^j(\boldsymbol{\lambda},\boldsymbol{\beta})}{\partial\boldsymbol{\lambda}} &= 0 \\
\sum_{i=1}^n(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j e^{(\alpha-\mathrm{I}[r_i<0])\mathbf{x}_j^T\boldsymbol{\lambda}+(\alpha-I_{ij}(\boldsymbol{\beta}_{\alpha_0}))\mathbf{x}_j^T\boldsymbol{\mu}} &= 0.
\end{aligned}$$

Denote by I_i indicator function of residuals

$$I_i = \mathrm{I}[r_i<0], \quad i=1,\dots,n.$$

Function $\mathbf{x}_j^T\boldsymbol{\lambda}$ of saddlepoint $\boldsymbol{\lambda}$ is a solution of the equation

$$\sum_{i=1}^n(\alpha-I_i)e^{(\alpha-I_i)\mathbf{x}_j^T\boldsymbol{\lambda}}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})} = 0$$

$$(1-\alpha)\sum_{i=1}^n I_i e^{(\alpha-1)\mathbf{x}_j^T\boldsymbol{\lambda}}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})} =$$

$$= \alpha\sum_{i=1}^n(1-I_i)e^{\alpha\mathbf{x}_j^T\boldsymbol{\lambda}}\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}$$

and is equal to

$$\mathbf{x}_j^T\boldsymbol{\lambda} = \log\left\{\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}I_i}{\sum_{i=1}^n\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}(1-I_i)}\right\}.$$

Denote

$$w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{ij}(\boldsymbol{\beta}_{\alpha_0})}}{\sum_{k=1}^n\left(\frac{\alpha}{1-\alpha}\frac{1-F_n^j(\boldsymbol{\beta}_{\alpha_0})}{F_n^j(\boldsymbol{\beta}_{\alpha_0})}\right)^{I_{kj}(\boldsymbol{\beta}_{\alpha_0})}}.$$

Hence we have

$$\mathbf{x}_j^T\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha) = \log\left\{\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_{ij}I_i}{\sum_{i=1}^n w_{ij}(1-I_i)}\right\}$$

and

$$\begin{aligned}
K^j(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha),\hat{\boldsymbol{\beta}}_\alpha) &= \log\sum_{l=1}^n\left(\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_{ij}I_i}{\sum_{i=1}^n w_{ij}(1-I_i)}\right)^{\alpha-I_l} w_{lj} \\
&= \log\left\{\left(\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_{ij}I_i}{\sum_{i=1}^n w_{ij}(1-I_i)}\right)^{\alpha}\times\right. \\
&\quad \left.\sum_{l=1}^n\left(\frac{\alpha}{1-\alpha}\frac{\sum_{i=1}^n w_{ij}(1-I_i)}{\sum_{i=1}^n w_{ij}I_i}w_{lj}I_l + w_{lj}(1-I_l)\right)\right\} \\
&= \log\left\{\left(\frac{1-\alpha}{\alpha}\frac{\sum_{i=1}^n w_{ij}I_i}{\sum_{i=1}^n w_{ij}(1-I_i)}\right)^{\alpha}\frac{\sum_{i=1}^n w_{ij}(1-I_i)}{1-\alpha}\right\} \\
&= \log\left\{\left(\frac{\sum_{i=1}^n w_{ij}I_i}{\alpha}\right)^{\alpha}\left(\frac{\sum_{i=1}^n w_{ij}(1-I_i)}{1-\alpha}\right)^{1-\alpha}\right\}.
\end{aligned}$$

Finally, the test statistic has the form

$$\begin{aligned}
2n\hat{h}(\hat{\boldsymbol{\beta}}_\alpha) =& -2nK(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha),\hat{\boldsymbol{\beta}}_\alpha) = -2\sum_{j=1}^n K^j(\boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_\alpha),\hat{\boldsymbol{\beta}}_\alpha) \\
=& -2\sum_{j=1}^n\log\left\{\left(\frac{\sum_{i=1}^n w_{ij}I_i}{\alpha}\right)^{\alpha}\left(\frac{\sum_{i=1}^n w_{ij}(1-I_i)}{1-\alpha}\right)^{1-\alpha}\right\}. \qquad (3.5)
\end{aligned}$$

□

Remark 7. *The asymptotic distribution of* (3.5) *has not been formally proved yet. Based on available results for saddlepoint tests, we suppose that the asymptotic distribution is* χ^2_{p+1} *(see Fig. 3.2). This distribution has been used in the simulations studies in this chapter and under hypothesis it provides very good approximation even for small sample sizes.*

3.3 Composite hypothesis

Suppose we now want to perform a test only on the first subvector of regression quantile:

$$H_0 : \boldsymbol{\beta}_{\alpha 1} = \boldsymbol{\beta}_{\alpha 1_0} \in \mathbb{R}^{p_1}, \qquad (3.6)$$

where $\boldsymbol{\beta}_\alpha = (\boldsymbol{\beta}_{\alpha 1}^T, \boldsymbol{\beta}_{\alpha 2}^T)^T$ and $\hat{\boldsymbol{\beta}}_\alpha = (\hat{\boldsymbol{\beta}}_{\alpha 1}^T, \hat{\boldsymbol{\beta}}_{\alpha 2}^T)^T$. Denote $\mathbf{x}_{i1} \in \mathbb{R}^{p_1}$ the subvector of $\mathbf{x}_i$ consisting of the first p_1 components of $\mathbf{x}_i$.

We will state theorems on the form of the test statistic and its asymptotic distribution under null hypothesis, but in the proofs only the derivation of the form of the test statistics will be outlined, as the asymptotic distribution can be proved using similar techniques already used in this chapter.

3.3.1 Parametric case

Theorem 11. *Test saddlepoint test statistic for composite hypothesis* (3.6) *has the form*

$$-2n \inf_{\boldsymbol{\beta}_{\alpha 2}} \log \int \left(\frac{G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \hat{\boldsymbol{\beta}}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma}\right)}{\alpha} \right)^{\alpha} \times$$

$$\left(\frac{1 - G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \hat{\boldsymbol{\beta}}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma}\right)}{1 - \alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i$$

and under hypothesis H_0

$$2nh(\hat{\boldsymbol{\beta}}_{\alpha 1}) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p1}.$$

Proof. Denote

$$\boldsymbol{\beta}^*_{\alpha} = (\boldsymbol{\beta}_{\alpha 1_0}, \boldsymbol{\beta}_{\alpha 2}).$$

Cumulant generating function of the score ψ is equal to

$$\begin{aligned} K_F(\boldsymbol{\lambda}, \boldsymbol{\beta}_{\alpha}) &= \log \mathsf{E}_F e^{\boldsymbol{\lambda}^T \psi(Y_i, \boldsymbol{\beta}_{\alpha})} \\ &= \log \int \left\{ e^{\alpha \boldsymbol{\lambda}^T \mathbf{x}_i} k(\mathbf{x}_i) \left(e^{-\boldsymbol{\lambda}^T \mathbf{x}_i} G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}^*_{\alpha})}{\sigma} \right) \right.\right. \\ &\quad \left.\left. + 1 - G\left(\frac{\mathbf{x}_i^T(\boldsymbol{\beta} - \boldsymbol{\beta}^*_{\alpha})}{\sigma} \right) \right) \right\} d\mathbf{x}_i. \end{aligned}$$

By solving

$$\frac{\partial K_F(\boldsymbol{\lambda}, (\boldsymbol{\beta}_{\alpha 1_0}, \boldsymbol{\beta}_{\alpha 2}))}{\partial \boldsymbol{\lambda}} = 0$$

we get the formula for $\boldsymbol{\lambda}^T \mathbf{x}_i$

$$\boldsymbol{\lambda}^T \mathbf{x}_i = \log \frac{1-\alpha}{\alpha} \frac{G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \boldsymbol{\beta}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma}\right)}{1 - G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \boldsymbol{\beta}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma}\right)}.$$

Therefore the test statistic can be rewritten as

$$h(\hat{\boldsymbol{\beta}}_{\alpha 1}) = \inf_{\boldsymbol{\beta}_{\alpha 2}} \log \int \left(\frac{G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \hat{\boldsymbol{\beta}}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma}\right)}{\alpha} \right)^{\alpha} \times$$

$$\left(\frac{1 - G\left(\frac{\mathbf{x}_{i1}^T(\boldsymbol{\beta}_{\alpha 1_0} - \hat{\boldsymbol{\beta}}_1) + \mathbf{x}_{i2}^T(\boldsymbol{\beta}_{\alpha 2} - \boldsymbol{\beta}_2)}{\sigma} \right)}{1 - \alpha} \right)^{1-\alpha} k(\mathbf{x}_i)\, d\mathbf{x}_i$$

and under hypothesis H_0

$$2nh(\hat{\boldsymbol{\beta}}_{\alpha 1}) \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p_1}.$$

□

3.3.2 Nonparametric case

Theorem 12. *The saddlepoint test statistic for composite hypothesis* (3.6) *equals*

$$-2nh(\hat{\boldsymbol{\beta}}_1) = -2 \inf_{\boldsymbol{\beta}_{\alpha 2}} \sum_{j=1}^n \log \left\{ \left(\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha} \right)^{\alpha} \left(\frac{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1 - \alpha} \right)^{1-\alpha} \right\}.$$

Proof. Denote

$$\begin{aligned} \boldsymbol{\beta}_{\alpha 2}^* &= \arg\min_{\boldsymbol{\beta}_{\alpha 2}} \left\{ -\log \left(\frac{1}{n} \sum_{i=1}^n \exp\left[\boldsymbol{\lambda}^T \psi(y_i, (\boldsymbol{\beta}_{\alpha 1_0}, \boldsymbol{\beta}_{\alpha 2})) \right] \right) \right\} \\ &= \arg\min_{\boldsymbol{\beta}_{\alpha 2}} \left\{ -\log \left(\frac{1}{n} \sum_{i=1}^n \exp\left[\boldsymbol{\lambda}^T (\alpha - I[y_i - x_{i1}^T \boldsymbol{\beta}_{\alpha 1_0} - x_{i2}^T \boldsymbol{\beta}_{\alpha 2} < 0] x_i \right) \right] \right\} \end{aligned}$$

and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_{\alpha 1_0}, \boldsymbol{\beta}_{\alpha 2})$. Then we follow the development for simple hypothesis, where

$$\mathbf{x}_j^T \boldsymbol{\mu} = \log \frac{1-\alpha}{\alpha} \frac{F_n^j(\boldsymbol{\beta}^*)}{1 - F_n^j(\boldsymbol{\beta}^*)}$$

solves

$$\sum_{i=1}^n \sum_{j=1}^n (\alpha - \mathrm{I}[r_i + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*)]) \mathbf{x}_j e^{(\alpha - \mathrm{I}[r_i + \mathbf{x}_j^T(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}^*)]) \mathbf{x}_j^T \boldsymbol{\mu}} = 0,$$

therefore the weights are of the form

$$w_{ij} = \frac{\left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}^*)}{F_n^j(\boldsymbol{\beta}^*)} \right)^{I_{ij}(\boldsymbol{\beta}^*)}}{\sum_{k=1}^n \left(\frac{\alpha}{1-\alpha} \frac{1 - F_n^j(\boldsymbol{\beta}^*)}{F_n^j(\boldsymbol{\beta}^*)} \right)^{I_{kj}(\boldsymbol{\beta}^*)}}$$

Hence

$$\mathbf{x}_j^T \boldsymbol{\lambda}(\hat{\beta}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2}) = \log \left\{ \frac{1-\alpha}{\alpha} \frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_2)}{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))} \right\},$$

where

$$I_i(\boldsymbol{\beta}_2) = I[y_i - x_{i1} \hat{\boldsymbol{\beta}}_{\alpha 1} - x_{i2} \boldsymbol{\beta}_{\alpha 2} < 0].$$

Then we can write

$$K^j((\hat{\boldsymbol{\beta}}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2}), \boldsymbol{\lambda}(\hat{\boldsymbol{\beta}}_{\alpha 1}, \boldsymbol{\beta}_{\alpha 2})) = \log \left\{ \left(\frac{\sum_{i=1}^n w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha} \right)^{\alpha} \left(\frac{\sum_{i=1}^n w_{ij}(1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1 - \alpha} \right)^{1-\alpha} \right\}$$

and

$$h(\hat{\boldsymbol{\beta}}_{\alpha 1}) = \inf_{\boldsymbol{\beta}_{\alpha 2}} -\frac{1}{n} \sum_{j=1}^{n} \log \left\{ \left(\frac{\sum_{i=1}^{n} w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha} \right)^{\alpha} \left(\frac{\sum_{i=1}^{n} w_{ij} (1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1-\alpha} \right)^{1-\alpha} \right\}.$$

and the test statistic is equal to

$$-2nh(\hat{\boldsymbol{\beta}}_1) = -2 \inf_{\boldsymbol{\beta}_{\alpha 2}} \sum_{j=1}^{n} \log \left\{ \left(\frac{\sum_{i=1}^{n} w_{ij} I_i(\boldsymbol{\beta}_{\alpha 2})}{\alpha} \right)^{\alpha} \left(\frac{\sum_{i=1}^{n} w_{ij} (1 - I_i(\boldsymbol{\beta}_{\alpha 2}))}{1-\alpha} \right)^{1-\alpha} \right\}.$$

□

3.3.3 Asymptotic distribution of saddlepoint test statistics

We performed simulation study in order to demonstrate the robustness of saddlepoint tests for regression quantiles in comparison to the Wald test. We performed 50000 simulations for various sample sizes both for parametric and nonparametric cases. The true parameter value for $\boldsymbol{\beta}$ in both cases is $(3, 2)^T$. Errors e_i were generated from normal, Laplace, logistic and contaminated normal distributions (with another normal distribution with larger variance). In parametric cases, as the distribution of observations was used normal standardized distribution. All tests were performed for regression quantiles with $\alpha = 0.25$. We considered one of the simplest situation where the i-th row of matrix $\boldsymbol{X}$ is equal to $(1, \frac{i-1}{n})$. For parametric case, we considered size samples $n = 5, 10, 20, 50, 100, 300, 1000, 10000$. As simulation study for nonparametric case was much more time consuming due to computations with matrices, we considered only $n = 21, 51, 101$.

In the nonparametric case, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}_\alpha$ used in the Wald test was estimated using the formula from Koenker and Bassett (1978) (implemented in the function `summary.rq` in `R`).

The results of the simulations are represented by plots (Fig. 3.1 and 3.2) and tables (Tab. 3.1 and 3.2). We plotted percentage of simulated test statistics that did not cross quantile of χ^2 (we used quantiles $\chi^2_{0.9}$, $\chi^2_{0.95}$ and $\chi^2_{0.99}$) against logarithm of number of observations. By stars we denoted results of saddlepoint tests (in tables denoted by "SAD"), by dots results of Wald test. We can see that for parametric case, the saddlepoint statistic and the Wald test behave very similarly. On the other hand, in nonparametric case saddlepoint test is much more precise and closer to χ^2 than the Wald test. These percentages are also written down in the tables.

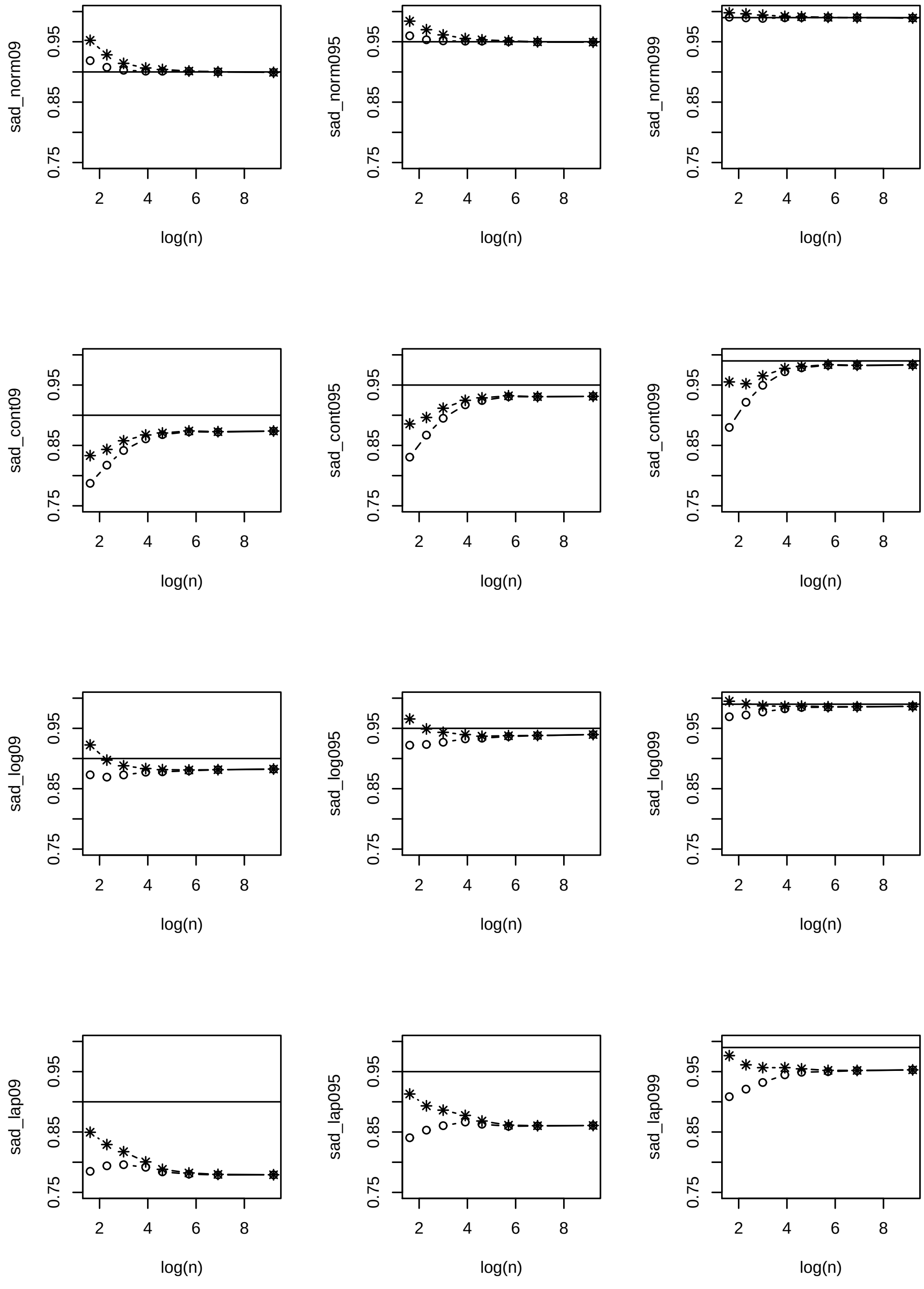


Figure 3.1: Comparison of parametric sadlepoint test and Wald test, probability that the test statistic does not exceed corresponding quantile of χ^2-distribution plotted against logarithm of number of observations

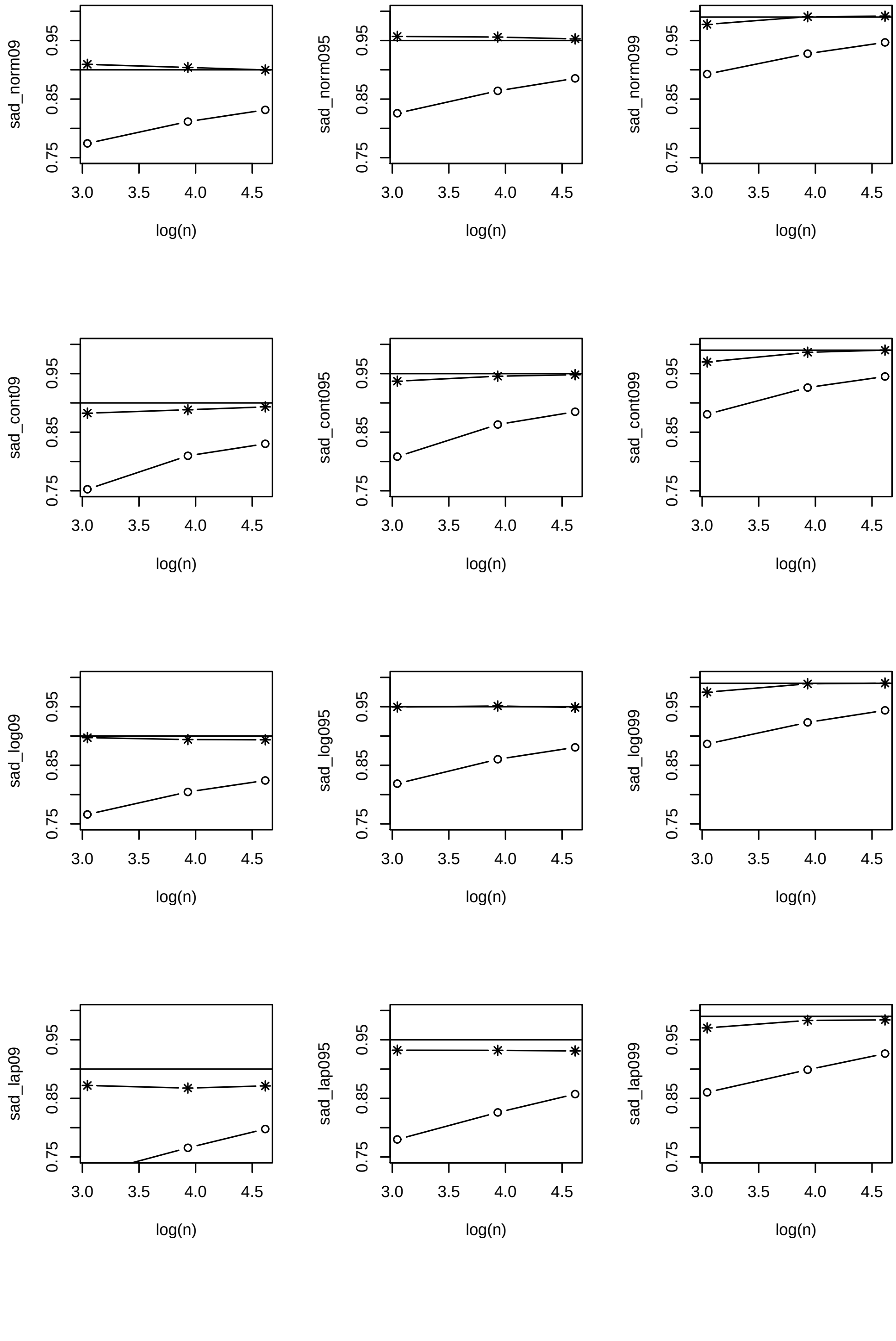


Figure 3.2: Comparison of nonparametric sadlepoint test and Wald test, probability that the test statistic does not exceed corresponding quantile of χ^2-distribution plotted against logarithm of number of observations

	$n = 21$			$n = 51$			$n = 101$		
	0.9	0.95	0.99	0.9	0.95	0.99	0.9	0.95	0.99
N(0, 1)									
SAD	0.9093	0.9570	0.9776	0.9040	0.9559	0.9907	0.8998	0.9527	0.9916
Wald	0.7744	0.8259	0.8927	0.8116	0.8641	0.9276	0.8316	0.8854	0.9468
cont									
SAD	0.8825	0.9371	0.9699	0.8883	0.9456	0.9864	0.8932	0.9481	0.9901
Wald	0.7526	0.8083	0.8806	0.8097	0.8631	0.9262	0.8302	0.8849	0.9451
Log									
SAD	0.8972	0.9495	0.9748	0.8939	0.9512	0.9891	0.8935	0.9487	0.9903
Wald	0.7661	0.8187	0.8864	0.8045	0.8603	0.9232	0.8241	0.8806	0.9438
Lap									
SAD	0.8720	0.9322	0.9703	0.8676	0.9320	0.9831	0.8713	0.9310	0.9839
Wald	0.7196	0.7799	0.8603	0.7656	0.8260	0.8989	0.7977	0.8573	0.9264

Table 3.1: Comparison of nonparametric saddlepoint test (SAD) and Wald test, H_0 : $\beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), contaminated N(0,.) with N(0,9) ($\varepsilon = 0.2$), Laplace and logistic distribution; probability that the test statistic does not exceed the corresponding quantile of χ^2-distribution

	$n=5$			$n=10$			$n=20$			$n=50$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
N(0, 1)												
SAD	0.9524	0.9842	0.9980	0.9284	0.9700	0.9963	0.9142	0.9614	0.9943	0.9066	0.9554	0.9921
Wald	0.9187	0.9599	0.9907	0.9077	0.9533	0.9896	0.9028	0.9515	0.9887	0.9013	0.9510	0.9898
cont												
SAD	0.8331	0.8856	0.9552	0.8434	0.8963	0.9522	0.8577	0.9117	0.9652	0.8674	0.9247	0.9776
Wald	0.7872	0.8306	0.8798	0.8173	0.8671	0.9215	0.8416	0.8950	0.9496	0.8609	0.9173	0.9721
Log												
SAD	0.9225	0.9655	0.9949	0.8974	0.9490	0.9903	0.8880	0.9434	0.9873	0.8834	0.9397	0.9864
Wald	0.8730	0.9221	0.9692	0.8692	0.9233	0.9721	0.8728	0.9270	0.9771	0.8774	0.9326	0.9824
Lap												
SAD	0.8496	0.9130	0.9765	0.8292	0.8933	0.9613	0.8174	0.8862	0.9566	0.8005	0.8775	0.9566
Wald	0.7849	0.8405	0.9085	0.7940	0.8531	0.9210	0.7959	0.8605	0.9320	0.7916	0.8665	0.9447

	$n=100$			$n=300$			$n=1000$			$n=10000$		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
N(0, 1)												
SAD	0.9041	0.9532	0.9915	0.9015	0.9515	0.9905	0.9002	0.9499	0.9900	0.8993	0.9493	0.9893
Wald	0.9012	0.9511	0.9903	0.9012	0.9504	0.9901	0.9003	0.9495	0.9900	0.8993	0.9494	0.9893
cont												
SAD	0.8706	0.9288	0.9807	0.8740	0.9322	0.9838	0.8728	0.9308	0.9828	0.8739	0.9313	0.9834
Wald	0.8679	0.9244	0.9786	0.8727	0.9309	0.9830	0.8721	0.9305	0.9822	0.8737	0.9314	0.9833
Log												
SAD	0.8816	0.9370	0.9868	0.8811	0.9378	0.9858	0.8815	0.9382	0.9858	0.8826	0.9397	0.9864
Wald	0.8780	0.9337	0.9847	0.8798	0.9366	0.9851	0.8814	0.9379	0.9856	0.8823	0.9397	0.9864
Lap												
SAD	0.7882	0.8682	0.9546	0.7822	0.8613	0.9521	0.7798	0.8604	0.9520	0.7792	0.8608	0.9529
Wald	0.7840	0.8629	0.9490	0.7803	0.8595	0.9501	0.7789	0.8601	0.9516	0.7791	0.8607	0.9529

Table 3.2: Comparison of parametric saddlepoint test (SAD) and Wald test, $H_0 : \beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.25$, N = 50000; data generated from: N(0,1), contaminated N(0,.) with N(0,9) ($\varepsilon = 0.2$), Laplace and logistic distribution

3.4 Simulation study

A numerical study comparing several tests with tests introduced in this chapter was carried out. We compared the proposed tests to the following alternatives:

- test based on the asymptotic distribution of regression quantiles,
- Wald test (with covariance matrix estimated using function `summary.rq` in `R`),
- likelihood ratio (LR)-type test,
- rank test.

The test statistic of LR type test has the form

$$LR_n = 2\frac{B}{A}\sum_{i=1}^{n}\left\{\rho_\alpha(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_{\alpha_0}) - \rho_\alpha(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_\alpha)\right\},$$

where $A = \mathsf{E}[\psi_\alpha^2] = \alpha(1-\alpha)$, $B = \mathsf{E}[\psi'_\alpha] = g(G^{-1}(\alpha))$ and under null hypothesis

$$LR_n \xrightarrow[n\to\infty]{\mathcal{D}} \chi^2_{p+1}.$$

B was estimated using the following relationship between the asymptotic variance of $\hat{\boldsymbol{\beta}}_\alpha$ and design matrix $\mathbf{X}$, A and B

$$\text{var}\hat{\boldsymbol{\beta}}_\alpha = \frac{A}{B^2}(\mathbf{X}^T\mathbf{X})^{-1}.$$

Test denoted by *asymp* is based on the asymptotic distribution of regression quantiles

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha\right) = \frac{1}{\sqrt{n}f(F^{-1}(\alpha))}Q_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i^T\psi_\alpha(E_{i\alpha}) + O_p(n^{-1/4}),$$

where $E_{i\alpha} = e_i - F^{-1}(\alpha)$ and

$$\psi_\alpha(x) = \begin{cases} \alpha & : x > 0 \\ \alpha - 1 & : x \le 0 \end{cases}$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_{p+1}\left(0, \mathbf{D}_0^{-1}\frac{\alpha(1-\alpha)}{f^2(F^{-1}(\alpha)}\right)$$

and the test statistic

$$n\frac{f^2(F^{-1}(\alpha))}{\alpha(1-\alpha)}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha_0})^T\mathbf{D}_0(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_{\alpha_0}) \sim \chi^2_{p+1},$$

where $\frac{1}{f(F^{-1}(\alpha))}$ was estimated by kernel estimator

$$\frac{\hat{\boldsymbol{\beta}}_{n1}(\alpha+\nu_n) - \hat{\boldsymbol{\beta}}_{n1}(\alpha-\nu_n)}{2\nu_n},$$

where ν_n was set to $\frac{3}{4}\alpha$.

The rank test is implemented in `R` in package `quantreg` using command `rq(..., se = "rank")`. This command returns confidence intervals described in Koenker (1994).

The results for saddlepoint test statistics are denoted by "SAD-par" for parametric saddlepoint test and "SAD-non" for nonparametric saddlepoint test.

We performed simulations for sample sizes $n = 21$, 31, 41 and 51, number of replications was set to 50000. The i-th row of matrix $\mathbf{X}$ was equal to $\mathbf{x}_i^T = (1, x_{i1} \ldots, x_{i5})$, where $x_{ij} \sim U(0,1), j = 1, 2, \ldots, 5$. The true value of parameter $\boldsymbol{\beta}$ was set to $\boldsymbol{\beta}^T = (3, 1, 2, 3, 4, 5)$. The errors $e_i, i = 1, \ldots, n$ were generated from two distributions with the same α-quantile: normal distribution $N(0, 1)$ and contaminated normal distribution $N(0, \cdot)$ with $N(0, 9)$ ($\varepsilon = 0.2$). The simulations were carried out for different values of α: 0.1, 0.15, 0.25 and 0.5

In the tables 3.3 - 3.6 we summarized the results of the numerical study under the hypothesis, the results are presented as a percentage (out of N) of simulated test statistics not exceeding the quantile (0.9, 0.95 and 0.99) of its asymptotic distribution. Notice that the approximation by χ^2-distribution is much more accurate for saddlepoint test statistics (both parametric and nonparametric) than for other compared tests. The accuracy of this approximation stays very high for parametric saddlepoint test even in contaminated model. These tests are very precise even for small values of α. The test based on the asymptotic distribution of averaged regression quantiles performes better than other classical tests. Notice that the results for rank test deteriorates for larger values of n. Although also other matrices $\mathbf{X}$ were considered, the results stayed similar.

In the tables 3.7 and 3.8 we summarized the results of a numerical study of performance of the compared tests under alternative. The percentages of rejected hypotheses are written down. We again considered the same sample sizes, values of α and distributions as in the previous study, but we considered different matrix $\mathbf{X}$. The i-th row of matrix $\mathbf{X}$ was equal to $\mathbf{x}_i^T = (1, x_{i1}, x_{i2})$, where $x_{ij} \sim U(0,1), j = 1, 2$. The percentage of rejected hypotheses is similar for all parametric tests, although it seems that saddlepoint test tends to reject the null hypothesis more often. Nonparametric saddlepoint test performes much better than the rank test - i.e. the only other nonparametric test we considered. Observe that the percentage of rejected hypotheses is larger for α further from 0.5.

$n = 21$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 21$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.86804	0.92902	0.98486	SAD - par.	0.88280	0.93950	0.98678
SAD - non.	0.88044	0.88634	0.88830	SAD - non.	0.94910	0.95918	0.96318
avg	0.63630	0.66554	0.70692	avg	0.74260	0.78260	0.83400
Wald	0.10420	0.14444	0.23148	Wald	0.28650	0.35050	0.47362
LR	0.31886	0.43328	0.63562	LR	0.50612	0.61812	0.78590
asymp	0.29802	0.31960	0.35678	asymp	0.46778	0.49726	0.54654
rank	0.86314	0.97908	0.99970	rank	0.70074	0.92168	0.99842

$n = 31$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 31$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.87644	0.93378	0.98454	SAD - par.	0.89132	0.94388	0.98770
SAD - non.	0.95298	0.95978	0.96240	SAD - non.	0.97114	0.98708	0.99320
avg	0.76566	0.80110	0.85028	avg	0.80814	0.84362	0.88538
Wald	0.20138	0.25206	0.35594	Wald	0.43304	0.49554	0.60172
LR	0.43262	0.53868	0.71466	LR	0.61814	0.70896	0.83724
asymp	0.49138	0.52192	0.57022	asymp	0.57320	0.60272	0.65146
rank	0.67218	0.87578	0.99606	rank	0.60458	0.80896	0.98270

$n = 41$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 41$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.88148	0.93776	0.98586	SAD - par.	0.89292	0.94542	0.98844
SAD - non.	0.97418	0.98258	0.98566	SAD - non.	0.97380	0.98434	0.99662
avg	0.82206	0.85528	0.89390	avg	0.84748	0.87820	0.91310
Wald	0.32152	0.38134	0.48658	Wald	0.53280	0.58982	0.68236
LR	0.52972	0.63186	0.78336	LR	0.68136	0.76238	0.86834
asymp	0.61246	0.63880	0.68164	asymp	0.65254	0.67888	0.72056
rank	0.61848	0.82932	0.98482	rank	0.58254	0.78580	0.97286

$n = 51$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 51$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.88536	0.93996	0.98666	SAD - par.	0.89586	0.94698	0.98914
SAD - non.	0.97366	0.98858	0.99494	SAD - non.	0.97446	0.99104	0.99698
avg	0.82450	0.86084	0.90184	avg	0.85108	0.88352	0.91784
Wald	0.42662	0.48592	0.58368	Wald	0.58198	0.63746	0.72374
LR	0.60806	0.69886	0.82726	LR	0.71506	0.78902	0.88868
asymp	0.63534	0.66484	0.70734	asymp	0.67636	0.70202	0.74096
rank	0.60102	0.80352	0.97936	rank	0.57236	0.77244	0.96780

Table 3.3: Comparison of accuracy of tests under $H_0 : \beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.1$ and $\alpha = 0.15$, N = 50000; data generated from: N(0,1)

$n = 21$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 21$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.90398	0.95206	0.98962	SAD - par.	0.91792	0.96042	0.99206
SAD - non.	0.94268	0.97126	0.98764	SAD - non.	0.66840	0.77144	0.86402
avg	0.80866	0.84476	0.88824	avg	0.85766	0.88724	0.92042
Wald	0.61122	0.67176	0.76364	Wald	0.82858	0.87172	0.92620
LR	0.73076	0.80806	0.90578	LR	0.85454	0.90960	0.96522
asymp	0.56892	0.59784	0.64588	asymp	0.67004	0.69622	0.73666
rank	0.58270	0.80776	0.99036	rank	0.53908	0.75868	0.97580
$n = 31$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 31$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.90588	0.95358	0.99036	SAD - par.	0.91516	0.95816	0.99202
SAD - non.	0.95312	0.98046	0.99592	SAD - non.	0.66496	0.78486	0.91788
avg	0.83580	0.86936	0.90506	avg	0.86430	0.89564	0.92762
Wald	0.64718	0.70142	0.78170	Wald	0.79630	0.84672	0.91344
LR	0.74792	0.81880	0.90770	LR	0.84310	0.90098	0.96322
asymp	0.62778	0.6554	0.69678	asymp	0.69028	0.71806	0.75908
rank	0.56548	0.77104	0.97078	rank	0.54516	0.75146	0.96300
$n = 41$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 41$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.90470	0.95292	0.98974	SAD - par.	0.91000	0.95642	0.99060
SAD - non.	0.95342	0.98018	0.99700	SAD - non.	0.64784	0.77602	0.92290
avg	0.85402	0.88566	0.91810	avg	0.86688	0.8964	0.9276
Wald	0.70236	0.7565	0.83576	Wald	0.78776	0.8396	0.9099
LR	0.79302	0.85934	0.93760	LR	0.84148	0.9011	0.96426
asymp	0.66898	0.69302	0.7333	asymp	0.70640	0.73192	0.77302
rank	0.56280	0.76130	0.96188	rank	0.5558	0.75356	0.95764
$n = 51$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 51$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.90548	0.95256	0.99034	SAD - par.	0.90852	0.95582	0.99156
SAD - non.	0.95232	0.9812	0.99782	SAD - non.	0.64336	0.77280	0.92098
avg	0.85818	0.88904	0.92328	avg	0.87742	0.90960	0.94244
Wald	0.7240	0.77910	0.85714	Wald	0.79668	0.85200	0.92038
LR	0.80724	0.87216	0.94580	LR	0.84676	0.90692	0.96862
asymp	0.69076	0.71664	0.75576	asymp	0.75904	0.78602	0.82538
rank	0.55464	0.75494	0.95762	rank	0.54912	0.74234	0.95080

Table 3.4: Comparison of accuracy of tests under $H_0 : \beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.25$ and $\alpha = 0.5$, N = 50000; data generated from: N(0,1)

$n = 21$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 21$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.52832	0.61044	0.73616	SAD - par.	0.60838	0.68272	0.79198
SAD - non.	0.70362	0.78770	0.86418	SAD - non.	0.81418	0.87836	0.93316
avg	0.68432	0.71002	0.74610	avg	0.77648	0.80996	0.85380
Wald	0.03148	0.05012	0.10260	Wald	0.18670	0.24184	0.35832
LR	0.23376	0.33026	0.52510	LR	0.45784	0.56820	0.74352
asymp	0.34594	0.36948	0.40810	asymp	0.52920	0.55738	0.60396
rank	0.84188	0.97074	0.99960	rank	0.69074	0.91630	0.99772

$n = 31$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 31$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.49938	0.58342	0.71614	SAD - par.	0.62172	0.69808	0.80124
SAD - non.	0.73562	0.81704	0.92078	SAD - non.	0.82706	0.89412	0.96670
avg	0.77958	0.81224	0.85368	avg	0.83172	0.86062	0.89768
Wald	0.07012	0.09888	0.17024	Wald	0.28308	0.34378	0.45642
LR	0.29808	0.39914	0.58788	LR	0.53440	0.63836	0.78894
asymp	0.55898	0.58638	0.63058	asymp	0.64688	0.67262	0.71232
rank	0.66538	0.86740	0.99500	rank	0.59874	0.80284	0.98268

$n = 41$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 41$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.48252	0.56840	0.69812	SAD - par.	0.62916	0.70622	0.81512
SAD - non.	0.72142	0.80280	0.91338	SAD - non.	0.82158	0.88094	0.96070
avg	0.84016	0.86798	0.90130	avg	0.87688	0.90066	0.92752
Wald	0.12516	0.16412	0.24976	Wald	0.37698	0.43874	0.54562
LR	0.36274	0.46618	0.64550	LR	0.60218	0.69642	0.82822
asymp	0.68836	0.71112	0.74684	asymp	0.74552	0.76472	0.79356
rank	0.61782	0.82756	0.98462	rank	0.58238	0.78702	0.97262

$n = 51$, $\alpha = 0.1$	0.9	0.95	0.99	$n = 51$, $\alpha = 0.15$	0.9	0.95	0.99
SAD - par.	0.47374	0.55988	0.69304	SAD - par.	0.64148	0.72206	0.82962
SAD - non.	0.71602	0.79900	0.91050	SAD - non.	0.83258	0.89372	0.96012
avg	0.86196	0.88778	0.91702	avg	0.89274	0.91414	0.93710
Wald	0.18766	0.23534	0.33052	Wald	0.44420	0.50370	0.60680
LR	0.43034	0.53270	0.70254	LR	0.64754	0.73154	0.85204
asymp	0.72784	0.74862	0.78144	asymp	0.77206	0.78944	0.81784
rank	0.59346	0.79878	0.97880	rank	0.57124	0.77018	0.96614

Table 3.5: Comparison of accuracy of tests under $H_0 : \beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.1$ and $\alpha = 0.15$, N = 50000; data generated from: contaminated N(0,.) with N(0,9) ($\varepsilon = 0.2$)

$n = 21$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 21$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.72562	0.79042	0.87354	SAD - par.	0.71024	0.79116	0.89402
SAD - non.	0.86140	0.91612	0.95962	SAD - non.	0.70332	0.79820	0.88620
avg	0.83910	0.86896	0.90304	avg	0.87988	0.90556	0.93214
Wald	0.55818	0.62516	0.72748	Wald	0.84880	0.8896	0.93632
LR	0.72508	0.80318	0.90290	LR	0.88092	0.92804	0.97384
asymp	0.64082	0.66730	0.70734	asymp	0.72046	0.74146	0.77482
rank	0.57848	0.80612	0.99062	rank	0.53954	0.75926	0.97612
$n = 31$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 31$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.76470	0.83204	0.91118	SAD - par.	0.72858	0.81406	0.91880
SAD - non.	0.88332	0.93580	0.98084	SAD - non.	0.69918	0.81444	0.93150
avg	0.86950	0.89430	0.92352	avg	0.8775	0.90428	0.93228
Wald	0.60114	0.66170	0.75316	Wald	0.81258	0.86138	0.92170
LR	0.74636	0.81888	0.90754	LR	0.86146	0.91548	0.97052
asymp	0.70808	0.73022	0.76412	asymp	0.71862	0.74146	0.77578
rank	0.56514	0.77072	0.96982	rank	0.54602	0.75380	0.96396
$n = 41$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 41$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.78388	0.84942	0.92720	SAD - par.	0.74074	0.82710	0.92812
SAD - non.	0.88812	0.93724	0.98332	SAD - non.	0.68256	0.80520	0.93660
avg	0.88938	0.91212	0.93664	avg	0.87472	0.90146	0.93078
Wald	0.66990	0.72860	0.81676	Wald	0.80260	0.85452	0.92052
LR	0.79272	0.86002	0.93728	LR	0.85538	0.91382	0.96978
asymp	0.75202	0.77154	0.80014	asymp	0.71238	0.73596	0.77268
rank	0.56060	0.76388	0.96354	rank	0.55346	0.75134	0.95660
$n = 51$, $\alpha = 0.25$	0.9	0.95	0.99	$n = 51$, $\alpha = 0.5$	0.9	0.95	0.99
SAD - par.	0.79906	0.86552	0.94150	SAD - par.	0.74558	0.83170	0.93486
SAD - non.	0.89326	0.94330	0.98456	SAD - non.	0.66450	0.79154	0.93506
avg	0.89170	0.91472	0.94018	avg	0.88350	0.91102	0.94010
Wald	0.69184	0.75050	0.83468	Wald	0.80770	0.85940	0.92502
LR	0.80168	0.86784	0.94262	LR	0.85996	0.91760	0.97284
asymp	0.76258	0.78204	0.81068	asymp	0.73088	0.75660	0.79492
rank	0.55466	0.75116	0.95632	rank	0.54434	0.74216	0.95154

Table 3.6: Comparison of accuracy of tests under $H_0 : \beta_\alpha = \beta + F^{-1}_{N(0,1)}(\alpha)$, $\alpha = 0.25$ and $\alpha = 0.5$, N = 50000; data generated from: contaminated N(0,.) with N(0,9) ($\varepsilon = 0.2$)

$\alpha = 0.1$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.89534	0.92984	0.97332	0.98994
SAD - non.	0.73094	0.61942	0.52058	0.76730
avg	0.49852	0.53468	0.65274	0.76812
Wald	0.93264	0.86132	0.86794	0.89246
LR	0.86746	0.81612	0.84390	0.88586
asymp	0.79022	0.66132	0.69950	0.73342
rank	0.16986	0.21478	0.31288	0.36860

$\alpha = 0.15$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.93020	0.95528	0.98672	0.99654
SAD - non.	0.57688	0.69702	0.68992	0.79936
avg	0.61382	0.64946	0.77426	0.88530
Wald	0.85406	0.82828	0.89502	0.94220
LR	0.81456	0.81826	0.90008	0.95244
asymp	0.77392	0.61636	0.73314	0.81408
rank	0.25696	0.28798	0.38790	0.44756

$\alpha = 0.25$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.95498	0.97672	0.99546	0.94000
SAD - non.	0.69776	0.77640	0.91566	0.99950
avg	0.75000	0.77996	0.92000	0.97334
Wald	0.83992	0.89362	0.95428	0.98180
LR	0.84472	0.91216	0.96846	0.98996
asymp	0.75088	0.72272	0.83446	0.91800
rank	0.31342	0.35198	0.46460	0.54018

$\alpha = 0.5$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.96418	0.98670	0.99846	0.99988
SAD - non.	0.85648	0.95336	0.99102	0.99862
avg	0.89870	0.93790	0.98906	0.99844
Wald	0.90090	0.95058	0.98858	0.99794
LR	0.94126	0.97554	0.99594	0.99950
asymp	0.74492	0.79934	0.92808	0.98280
rank	0.34978	0.40204	0.52424	0.61698

Table 3.7: Percentage of rejected hypotheses at significance level 0.05, normal distribution, hypothetical value $\boldsymbol{\beta}_{\alpha_0} = (3+F^{-1}(\alpha), 0, 0)$, real value $(3+F^{-1}(\alpha), 1, 1)$

$\alpha = 0.1$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.90232	0.9271	0.94694	0.96056
SAD - non.	0.40086	0.25676	0.18138	0.31622
avg	0.32664	0.27944	0.28244	0.31106
Wald	0.90910	0.85158	0.82626	0.81212
LR	0.77278	0.72556	0.70558	0.71012
asymp	0.58904	0.46606	0.43816	0.41598
rank	0.14552	0.17772	0.23228	0.25988

$\alpha = 0.15$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.91756	0.93794	0.96718	0.98088
SAD - non.	0.31484	0.35970	0.33162	0.42580
avg	0.39644	0.33128	0.39394	0.48732
Wald	0.83682	0.79736	0.82634	0.87516
LR	0.73366	0.71620	0.76536	0.82802
asymp	0.57142	0.46296	0.48540	0.51524
rank	0.21062	0.24036	0.30880	0.35352

$\alpha = 0.25$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.94094	0.96584	0.98914	0.99678
SAD - non.	0.50066	0.54836	0.73920	0.76134
avg	0.51160	0.52012	0.66588	0.7618
Wald	0.80398	0.84622	0.90860	0.95100
LR	0.76582	0.83392	0.90780	0.94998
asymp	0.58312	0.5720	0.63112	0.68688
rank	0.27856	0.31986	0.42128	0.49090

$\alpha = 0.5$	$n = 21$	$n = 31$	$n = 41$	$n = 51$
SAD - par.	0.94912	0.97446	0.99486	0.99934
SAD - non.	0.69662	0.85218	0.94766	0.98370
avg	0.69274	0.77808	0.92720	0.98020
Wald	0.78084	0.86438	0.94938	0.98458
LR	0.83438	0.90558	0.97136	0.99260
asymp	0.54622	0.61292	0.76170	0.88856
rank	0.28530	0.33158	0.42996	0.51590

Table 3.8: Percentage of rejected hypotheses at significance level 0.05, contaminated normal distribution, hypothetical value $\boldsymbol{\beta}_{\alpha_0} = (3 + F^{-1}(\alpha), 0, 0)$, real value $(3 + F^{-1}(\alpha), 1, 1)$

Chapter 4

Saddlepoint approximation for the density based on I-divergence

In this chapter we will not work with classical setup of n i.i.d. random variables and will consider n independent but not necessarily identically distributed random variables instead. Let us assume $Y_1, \ldots, Y_n$ are independent and the distribution of Y_i depends on the parameter γ_i. We will specifically deal with Y_i from the exponential family. The form of pdf of exponential family will allow us to write the saddlepoint approximation to the density of sufficient statistic and MLE in the explicit form. The inference will be based on a special case of Kullback-Leibler divergence, called I-divergence. Part of this chapter was submitted, see Sabolová et al. (2014).

4.1 Information theory, I-divergence

In order to connect the information theory with the statistical decision theory we study the measures of difference between statistical models used in information theory. These are usually referred to as information divergences and their development started after the introduction of information entropy in (Shannon, 1948). Entropy is commonly viewed as the measure of the amount of uncertainty contained in a random variable Y based on its pdf (continuous case) or pmf (discrete case). The entropy of a continuous random variable Y having pdf $f(y)$ with respect to Lebesgue measure is defined as

$$H(Y) = -\int_Y f(y) \log f(y) dx. \tag{4.1}$$

In this chapter we concentrate on the Kullback-Leibler (KL) divergence. This divergence was introduced as a generalization of Shannon's entropy in (Kullback and Leibler, 1951). Its "creation" was initiated by a need to solve the statistical problems of discrimination while satisfying the additional criterion – the chosen statistic has to contain all relevant information provided by the sample. The KL-divergence is defined as the mean information in observation y (on the state space Y) for discrimination between f_1 and f_2, where these stand for pdfs w.r.t.

probability measures μ_i, $i = 1, 2$.

$$D_{KL}(f_1, f_2) = \int_Y f_1(y) \log \frac{f_1(y)}{f_2(y)} dy. \tag{4.2}$$

Remark 8. *Symmetry and triangle inequality are not satisfied for KL-divergence, thus it is not a distance. Notice positive semidefinitness, i.e. $D_{KL}(f_1, f_2) \geq 0$ with equality only if $f_1(y) = f_2(y)$ a.e. (see (Kullback and Leibler, 1951)).*

We will specifically deal with the exponential family that includes many common distributions and its pdf has a convenient form that will enable us to rewrite KL-divergence in a form suitable for consequent computations (i.e. derivation of saddlepoint approximation to the density of sufficient statistic and MLE).

Definition 2. *Probability density function for the exponential family has the form*

$$h(y_i|\gamma_i) = \exp\{\log f(y_i) + \gamma_i T(y_i) + \log g(\gamma_i)\} \quad i = 1, \ldots, n. \tag{4.3}$$

Joint density for $y_1, \ldots, y_n$ has the form

$$\begin{aligned} h(y|\gamma) &= \prod_{i=1}^{n} f(y_i) g(\gamma_i) \exp\{\gamma_i T(y_i)\} \\ &= \exp\left\{\sum_i \log(f(y_i)) + \sum_i \gamma_i T(y_i) + \sum_i \log(g(\gamma_i))\right\} \\ &= \exp\left\{-\psi(y) + \boldsymbol{t}(y)^T \boldsymbol{\gamma} - \kappa(\gamma)\right\}, \end{aligned} \tag{4.4}$$

where

$$\begin{aligned} \psi(y) &= -\sum_i \log(f(y_i)), \\ \kappa(\gamma) &= -\sum_i \log(g(\gamma_i)). \end{aligned} \tag{4.5}$$

In the following, we will work with the setup used in (Pázman, 1993) where certain smoothness and regularity assumptions on the exponential family have been considered, see p. 217–218. In particular, we will suppose that

$$\{t(y) : y \in Y\} \subseteq \{\mathrm{E}_\gamma[t(y)] : \gamma \in \Gamma\}$$

and

$$\frac{\partial \kappa(\gamma)}{\partial \gamma} = \mathrm{E}_\gamma[t(y)]$$

(see Barndorf-Nielsen (1978)). These two properties together with (4.5) imply

$$\frac{\partial \kappa(\gamma)}{\partial \gamma_i}|_{\gamma=\hat{\gamma}_y} = -\frac{g'_{\gamma_i}(\hat{\gamma}_i)}{g(\hat{\gamma}_i)} = T(y_i) \tag{4.6}$$

where $\hat{\gamma}_i$ is a MLE of the parameter γ_i and $\hat{\gamma}_y = (\hat{\gamma}_1, \ldots, \hat{\gamma}_n)$ (see Stehlík (2003)).

Let us now introduce a special case of KL-divergence based on (4.6).

Definition 3. *The I-divergence of $h(y|\gamma)$ from $h(y|\gamma^*)$ has the form*

$$I_n(\gamma^*,\gamma) = \sum_i -\frac{g'_{\gamma_i}(\hat{\gamma}_i)}{g(\gamma_i^*)}(\gamma_i^* - \gamma_i) + \sum_i \log\left(\frac{g(\gamma_i^*)}{g(\gamma_i)}\right) \tag{4.7}$$

and I-divergence of the canonical parameter γ from sampled vector y has the form

$$I_n(y,\gamma) = I(\hat{\gamma}_y,\gamma) = \sum_i (T(y_i)\hat{\gamma}_i + \log g(\hat{\gamma}_i)) - \sum_i (T(y_i)\gamma_i + \log g(\gamma_i)). \tag{4.8}$$

4.2 Testing hypotheses based on I-divergence

I-divergence in the exponential family can be rewritten as a sum of two random variables, each one representing a test statistic.

$$\begin{aligned}
\mathrm{I}_n(y,\gamma) =& \sum_i (T(y_i)\hat{\gamma}_i + \log g(\hat{\gamma}_i)) - \sum_i (T(y_i)\gamma_i + \log g(\gamma_i)) \\
=& \sum_i (T(y_i)\hat{\gamma}_i + \log g(\hat{\gamma}_i)) - \sum_i (T(y_i)\gamma_i + \log g(\gamma_i)) \\
& \pm \sum_i (T(y_i)\hat{\gamma}_{MLE} + \log g(\hat{\gamma}_{MLE})) \\
=& \left[\sum_i (T(y_i)\hat{\gamma}_{MLE} + \log g(\hat{\gamma}_{MLE})) - \sum_i (T(y_i)\gamma_i + \log g(\gamma_i))\right] \\
& + \left[\sum_i (T(y_i)\hat{\gamma}_i + \log g(\hat{\gamma}_i)) - \sum_i (T(y_i)\hat{\gamma}_{MLE} + \log g(\hat{\gamma}_{MLE}))\right] \\
=& \mathrm{R}_n + \mathrm{S}_n
\end{aligned}$$

where $\hat{\gamma}_{MLE}$ is the MLE of the parameter γ based on the whole observed sample $y_1, ldots, y_n$.

S_n represents test statistics of a likelihood ratio test for

$$H_0 : \gamma_1 = \ldots = \gamma_N \quad \text{vs.} \quad H_1 : \text{ arbitrary suitable } \gamma, \quad \gamma \in \Gamma.$$

R_n represents test statistics of a likelihood ratio test for

$$H_0 : \gamma_1 = \ldots = \gamma_N = \gamma_0 \quad \text{vs.} \quad H_1 : \gamma \neq \gamma_0.$$

4.2.1 Independence of R_n and S_n

Independence of R_n and S_n for gamma-distributed random variables was proved in Stehlík (2003). Pareto-distributed random variable with known threshold x_m belongs to the exponential family (unlike Pareto-distributed random variable with unknown x_m), its pdf being

$$f_Y(y) = \alpha\frac{x_m^\alpha}{y^{\alpha+1}}, \quad x_m > 0, \quad \alpha > 0. \tag{4.9}$$

Now let us consider the twodimensional sample from Pareto distribution and the transformation

$$\log \frac{y_1}{x_m} = ts, \qquad \log \frac{y_2}{x_m} = t(1-s),$$

where $t \in (0, \infty)$, $s \in (0, 1)$, having Jacobian equal to

$$|J| = x_m^2 t e^t.$$

Thus, the density of the transformed vector has form

$$g_{S,T}(s,t) = \alpha_1 \alpha_2 t e^{-\alpha_1 ts} e^{-\alpha_2 t(1-s)}.$$

We obtain radial and spherical components of the form

$$\begin{aligned} R_{2H_0} &= -2 + 2\log 2 - 2\log t + (-\gamma_0 - 1)t \\ S_2 &= -\log s - \log(1-s) - 2\log 2 \end{aligned}$$

Thus we have independence of R_n and S_n.

4.2.2 Graphical method for testing $H_0 : \gamma = \gamma_0$ in the exponential distribution

Suppose that $\{Y_i\}_{i=1}^n$ is a random sample from exponential distribution with rate γ and corresponding pdf

$$h(y_i|\gamma) = \gamma e^{-\gamma y_i}, \quad y_i \geq 0.$$

For testing the hypothesis about the unknown parameter $\gamma = (\gamma_1, \ldots, \gamma_n)^T$

$$H_0 : \gamma = \gamma_{\mathbf{0}} \quad (H_1 : \gamma \neq \gamma_{\mathbf{0}})$$

based on available observations $\mathbf{y} = (y_1, \ldots, y_n)^T$ we exploit the I-divergence introduced in Stehlík (2003):

$$I_n(\mathbf{y}, \gamma_{\mathbf{0}}) = n \ln \sum_i y_i - \sum_i \ln y_i + \gamma_0 \sum_i y_i - n \ln(\gamma_0 \sum_i y_i) - n.$$

For the exponential case, the exact form of cdf for I_n, $n \leq 4$ is derived in (Stehlík et al., 2014). For $n = 1$ and $y \sim Exp(\gamma)$ the cdf of $I_1(\hat{\gamma}, \gamma)$ is equal to

$$F_1(x) = \begin{cases} \exp\{\hat{\gamma}^{-1}\gamma W_0(-\exp\{-1-x\}\} - \exp\{\hat{\gamma}^{-1}\gamma W_{-1}(-\exp\{-1-x\}\}, & x > 0 \\ 0 & x \leq 0, \end{cases}$$

where W_0, W_{-1} are the two real-valued branches of Lambert-W function. Quantile plot of $F_1(I_1)$ under $H_0 : \gamma = 1$ against quantiles of uniform distribution U(0,1) is displayed in Fig. 4.1. In this example we generated $M = 1000$ random samples each consisting of 1 random variable from the distribution $Exp(1)$. The plotted points form the diagonal in the plot, thus we do not reject the hypothesis.

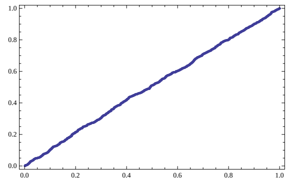


Figure 4.1: Simulations for exponential case with exact F_1, number of samples 1000 and $\gamma = 1$

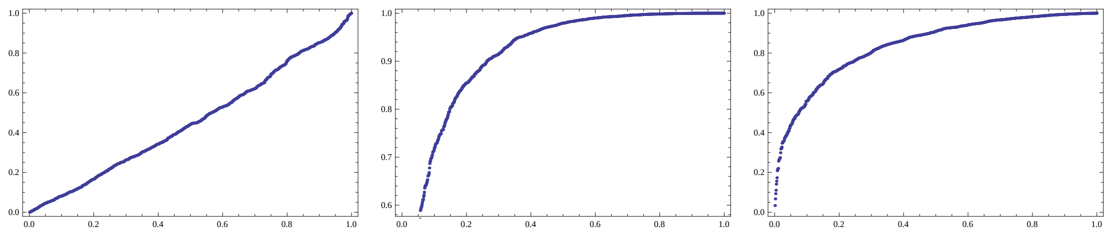
As formulas for $n \geq 2$ are rather complicated, we exploit the deconvolution of I_n to two independent parts R_n and S_n as follows:

$$I_n(y, \gamma_0) = R_n + S_n$$

$$R_n = \gamma_0 \sum_i y_i - n \ln(\gamma_0 \sum_i y_i) - n + n \ln n$$
$$S_n = n \ln \sum_i y_i - n \ln n - \sum_i \ln y_i.$$

To construct the cdf of I_n and thus the quantile plot we use asymptotic distribution of R_n and S_n. According to (Stehlík, 2003), R_n is asymptotically χ^2_1-distributed. Asymptotic distribution of S_n was derived in (Bartlett and Kendall, 1946) and is equal to

$$\frac{1}{2}\left(1 + \frac{1 + \frac{1}{n}}{6}\right)\chi^2_{n-1}.$$



(a) $\gamma = 1$, $\gamma_0 = 1$ (b) $\gamma = 1$, $\gamma_0 = 2$ (c) $\gamma = 1$, $\gamma_0 = \frac{1}{2}$

Figure 4.2: Simulations for exponential case for 1000 samples with $n = 50$.

Quantile plots of $\hat{F}_n(I_n)$ for $\hat{F}_n$ being the asymptotic cdf of I_n for $M = 1000$ samples of size $n = 50$ from $Exp(1)$ against quantiles of uniform distribution U(0,1) for different hypotheses $H_0 : \gamma = \gamma_0$ can be found in Fig. 4.2a, 4.2b and 4.2c.

In case when the data were drawn from $Exp(1)$ and the hypothetical value for the unknown parameter γ was set to 1, the plotted points form a diagonal in the quantile plot for asymptotic cdf of I_n against cdf of uniform distribution $U(0, 1)$. When testing the hypothesis that the unknown parameter γ equals 2 or 1/2 (see Fig. 4.2b and 4.2c), the results in the plots differ significantly from the straight line forming diagonal. Thus we reject hypotheses in both cases.

4.3 Saddlepoint approximations for linear exponential family

The derivation of saddlepoint approximation for density of a sufficient statistic and density of MLE will be based on theory introduced in Pázman (1993).

Theorem 13. *(Saddlepoint approximation of the density of a sufficient statistics T) The saddlepoint approximation of the density of the vector of a sufficient statistic T is equal to*

$$q_T(t|\gamma) = (2\pi)^{-n/2} \left(\prod_i \frac{g'(\hat{\gamma}_i)^2 - g''(\hat{\gamma}_i) g(\hat{\gamma}_i)}{g^2(\hat{\gamma}_i)} \right)^{-1/2} \prod_i \frac{g(\gamma_i)}{g(\hat{\gamma}_i)} \exp\{t_i[\gamma_i - \hat{\gamma}_i]\}. \tag{4.10}$$

and the saddlepoint approximation of the density of MLE $\hat{\gamma}$ is equal to

$$q_{MLE}(\hat{\gamma}|\gamma) = (2\pi)^{-n/2} \left(\prod_i \frac{g'(\hat{\gamma}_i)^2 - g''(\hat{\gamma}_i) g(\hat{\gamma}_i)}{g^2(\hat{\gamma}_i)} \right)^{1/2} \prod_i \frac{g(\gamma_i)}{g(\hat{\gamma}_i)} \exp\{t_i[\gamma_i - \hat{\gamma}_i]\}. \tag{4.11}$$

Proof. The saddlepoint approximation of the density of the vector T is equal to (see (Pázman, 1993), definition 9.3.1):

$$q_T(t|\gamma) = (2\pi)^{-k/2} |\Sigma_{\hat{\gamma}(t)}|^{-1/2} \exp\{-I_n(\hat{\gamma}(t), \gamma)\},$$

where

$$\Sigma_\gamma = \frac{\partial^2 \kappa}{\partial \gamma \partial \gamma^T},$$

$I_n(.,.)$ is given in (4.8) and $\hat{\gamma}(t)$ is a solution of maximization problem

$$\hat{\gamma}(t) = \arg\max_{\gamma \in \Gamma} [t^T \gamma - \kappa(\gamma)].$$

$$\frac{\partial^2 \kappa}{\partial \gamma_i \partial \gamma_j} = \frac{1}{g^2(\gamma_i)} \left(g'(\gamma_i) g'(\gamma_j) - g''(\gamma_i) g(\gamma_j) \, \mathrm{I}_{[i=j]} \right)$$

Matrix Σ_γ is diagonal and equals

$$\frac{\partial^2 \kappa}{\partial \gamma \partial \gamma^T} = diag \left\{ \frac{g'(\gamma_i)^2 - g''(\gamma_i) g(\gamma_i)}{g^2(\gamma_i)} \right\}$$

and its determinant therefore can be rewritten as a product of its diagonal elements

$$\left| \frac{\partial^2 \kappa}{\partial \gamma \partial \gamma^T} \right| = \prod_i \frac{g'(\gamma_i)^2 - g''(\gamma_i) g(\gamma_i)}{g^2(\gamma_i)}$$

The other part of the saddlepoint approximation of the density can be expressed in a following way

$$
\begin{aligned}
\exp\{-I_n(\hat{\gamma}(t),\gamma)\} &= \exp\left\{\sum_i (t_i\gamma_i + \log g(\gamma_i)) - \sum_i (t_i\hat{\gamma}_i + \log g(\hat{\gamma}_i))\right\} \\
&= \frac{\prod_i \exp\{t_i\gamma_i\} g(\gamma_i)}{\prod_i \exp\{t_i\hat{\gamma}_i\} g(\hat{\gamma}_i)} \\
&= \prod_i \frac{g(\gamma_i)}{g(\hat{\gamma}_i)} \exp\{t_i[\gamma_i - \hat{\gamma}_i]\}
\end{aligned}
$$

Finally we get the saddlepoint approximation of the density of sufficient statistic in exponential family of the form (4.10).

Formula for saddlepoint approximation of the density of MLE follows from

$$
q(\hat{\gamma},\gamma) = det\,(\Sigma_{\hat{\gamma}})\, q_T(t,\gamma)|_{t=t(\hat{\gamma})},
$$

see (Pázman, 1993). □

4.4 Application to real data - methane emissions

Methane is a product of anaerobic decomposition processes of the organic matter cycles mainly in water-saturated soils. These processes require the synergy or syntrophic cooperation between anaerobic bacteria and methanogenic archaea. Organic matter is firstly hydrolysed and fermented, and the products that are formed are the compounds used for methanogenesis (Le Mer and Roger, 2001). Methane production can be described as dissipative process of entropy in which highly organized organic structures are decomposed to basic simple compounds. The process of releasing the methane from soil and plant stand is highly determined by its production. Methane production depends on the occurrence of methanogenic bacteria, the amount of decomposed organic matter and suitable anaerobic conditions. Methane releasing from deeper soil layers is influenced by many processes, including methane oxidation by the methanetrophic bacteria (see Shukla et al. (2013)) and also by its means of being released from the soil environment.

Methane emissions are typically modelled via trend process fitting, which makes the amount of stochasticity and chaos present in the system questionable. In Jordanova et al. (2013b), this dependence has been modelled by a time series model. The trend component has been estimated by the Ordinary Least Squares technique. The noise component is represented by sum of an infinite moving average model with Pareto-like positive and negative parts of the innovations and independent identically distributed (i.i.d.) innovations with similar tail behaviour. Pareto tails have been also justified by robust tests for normality against Pareto tails (see Stehlík et al. (2013)). Such moving average time series could be considered as born by a point process, which is not homogeneous (see Jordanova et al. (2013a)). The process of methane release from soil is both chaotic and stochastic. In Jordanova et al. (2013b) the relation between stochastic and chaotic model was outlined: The parameters typically associated with chaos (both deterministic and stochastic) are measures of dimension, rate of information generated (entropy), and the Lyapunov spectrum. Entropies, as a measure of self-organization (or level of chaos of the system) in our case correspond to heavy tail parameters (in our stochastic model these are Pareto tails of lower and upper exceedances over thresholds, respectively). In Jordanova et al. (2013b), large tail parameters of the underlying stochastic process for both upper and lower exceedances were observed. This confirms that the amount of chaos measured through correlation dimension or entropy is not large.

As we have already mentioned the Pareto tails are proved to be suitable for modelling the release of methane by bacteria. To measure the amount of unpredictability or chaos involved in such system we exploit the information entropy (4.1). If we assume strict parametric assumption of the Pareto distribution $Pareto(x_m, \alpha)$ with scale parameter x_m and shape parameter α having the probability density function (pdf) (4.9) then we can derive the theoretical relationship between the tail parameter and its entropy (see Yari and Borzadaran (2010)):

$$H(Y) = \log\left(\frac{x_m}{\alpha}\right) + \frac{1}{\alpha} + 1. \tag{4.12}$$

This is always decreasing function for increasing α independent of the value of x_m:

$$\frac{\partial H(Y)}{\partial \alpha} = -\frac{1}{\alpha} - \frac{1}{\alpha^2} < 0 \quad (\alpha > 0).$$

Jordanova et al. (2013b) observed that large values of tail parameters α confirming moderate level of chaos in the system, justifying underlying biochemical intuition.

For real data on methane emissions the value of threshold x_m is usually not provided explicitly. We shall treat x_m as a nuisance parameter, which is however substantial for interpretation since it distinguishes between normal diffusion and anomalous diffusion (chaos). Our conjecture for the source of the chaos are various interactions (e.g. way of escaping of methane). The process of diffusion and other non-specific ways of methane releasing (e.g. ebullition) occur simultaneously. Normal diffusion relates to stochasticity, non-specific ways of releasing of methane relate to chaotical behavior of the system.

Both parts, stochastic and chaotic are hardly separable. Therefore the simulated process compared via I-divergence (4.8) to real data process can be taken only in upper tails of real data. Thus peaks of methane emissions can correspond to various spontaneous releases of methane which we understand to be chaotic in its nature.

Real methane data

We will analyze the residuals $Z, Z-, Z+$ of methane emissions taken from infinite moving average model (8) in Jordanova et al. (2013b), where only time is taken as a regressor. We trimmed the original data sets with sample sizes $M_{Z+} = 998$, $M_{Z-} = 971$ by 30%, thus we obtained Pareto distributed samples of sample sizes M_+, M_-.

Before we proceed to testing the null hypothesis about the parameter $\gamma = (\gamma_1, \ldots, \gamma_n)$ it is necessary to deal with the unknown nuisance parameter x_m. To obtain the desired value of x_m we exploit a maximum entropy principle (Shore and Johnson (1980), Jaynes (1982)). This principle states that from the set of all possible distributions, now represented by a set of Pareto distributions with fixed α and unknown x_m, we should choose the distribution with the highest entropy. For fixed γ, ($\gamma = -\alpha - 1$), and x_m on the interval $(0, y_{min})$, where y_{min} is the minimal value from the sample $y_1, \ldots, y_n$, the entropy is an increasing function of x_m, see (4.12). Thus, the value of x_m chosen for the proposed graphical method should be close to the minimal value of the sample $y_1, \ldots, y_n$. Since data are Pareto distributed, $x_m < y_i \quad \forall i$, we set $x_m = 0.99 \times y_{min}$.

Let us still assume the shape parameter α is known. Thus the uncertainty of the current system, represented by entropy (4.12), now depends on the value of the scale parameter x_m. If the chemometrician is expecting the system to be more deterministic, one should choose the value of x_m close to zero. This coincides with the fact that then the exceedances over threshold x_m are assigned with lower probability. On the other hand, if the system can be more chaotic, that is the exceedances over the threshold x_m occur with high probability, the value of x_m should be close to the minimal value of the sample $y_1, \ldots, y_n$. For our Pareto data $Z+$ with $\alpha = 1.3$ and the minimal value of the sample 2.05×10^{-6} the relation between entropy and value of x_m is shown in Fig.4.3. For lower values of entropy the value of x_m is really close to zero, the maximum achievable entropy is obtained for x_m close to y_{min}.

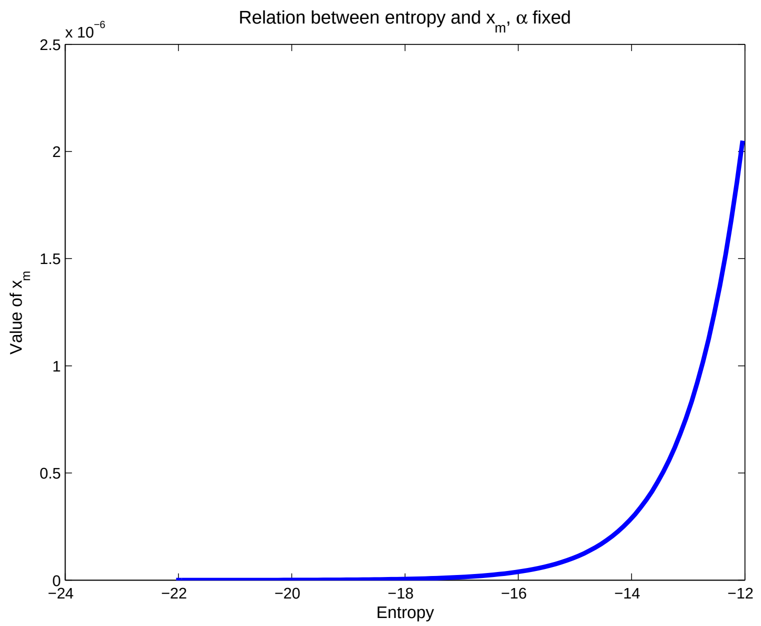
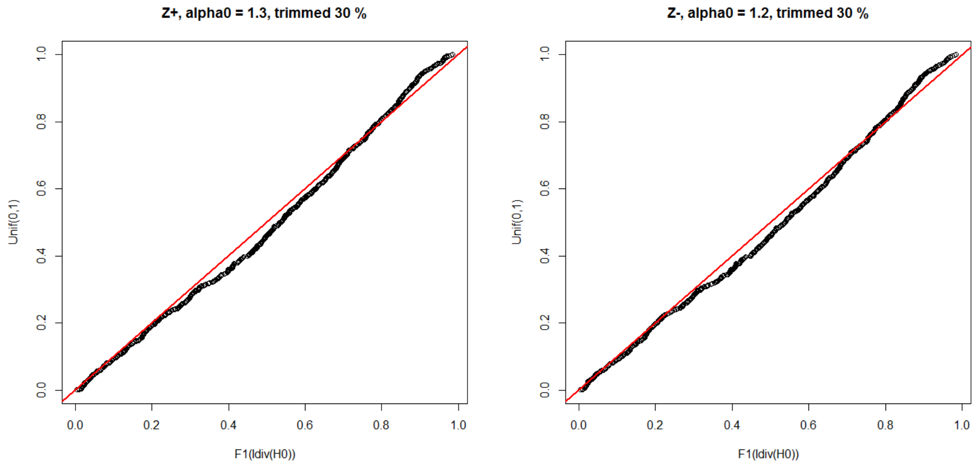


Figure 4.3: The relation between entropy and scale parameter x_m, fixed parameter α.

For chosen hypotheses $H_0 : \gamma = \gamma_0$ we transformed the both data samples using the relationship between Pareto and exponential distribution and used the procedure described in Section 4.2.2.

Based on expert's opinion, we set $\alpha_0 = 1.3$ for trimmed data set $Z+$ and $\alpha_0 = 1.2$ for trimmed data set $Z-$. Proposed graphical method confirms these hypotheses as we can see in Fig. 4.4a and Fig. 4.4b.



(a) Pareto methane data - Z+ (b) Pareto methane data - Z-

Figure 4.4: Pareto real data case: $\gamma_0 = 1.3$ for Z+, $\gamma_0 = 1.2$ for Z-.

In order to derive a saddlepoint approximation it is first necessary to compute density of a sufficient statistic t, that for Pareto distribution has the form $t = \log y$. Using results from Pázman (1993) we get the saddlepoint approximation for the density of t

$$q_T(t|\gamma) = (2\pi)^{-n/2} \prod_i (-\gamma_i - 1) x_m^{-\gamma_i - 1 + \frac{1}{\log x_m - t_i}} \exp\left\{ t_i \left(\gamma_i - \frac{1}{\log x_m - t_i} + 1 \right) \right\}.$$

Saddlepoint aproximation for the density of MLE is based on the density of the sufficient statistic (see Pázman (1993))

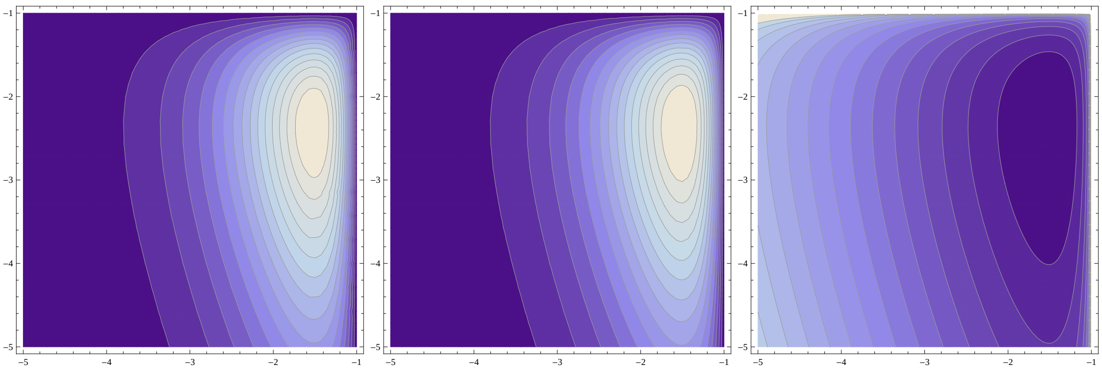
$$\hat{\gamma}_i = \frac{1}{\log x_m - t_i} - 1, \qquad i = 1, \ldots, n$$

and its formula reads

$$\begin{aligned} q_{MLE}(\hat{\gamma}|\gamma) =& (2\pi)^{-n/2} \prod_i (\log x_m - t_i)^2 (-\gamma_i - 1) x_m^{-\gamma_i - 1 + \frac{1}{\log x_m - t_i}} \\ & \times \exp\left\{ t_i \left(\gamma_i - \frac{1}{\log x_m - t_i} + 1 \right) \right\} |_{t=t(\hat{\gamma})} \\ =& (2\pi)^{-n/2} \prod_i \frac{-\gamma_i - 1}{(\hat{\gamma}_i + 1)^2} x_m^{-\gamma_i + \hat{\gamma}_i} \exp\left\{ (\gamma_i - \hat{\gamma}_i) \left(\log x_m - \frac{1}{\hat{\gamma}_i + 1} \right) \right\}. \end{aligned}$$

In order to illustrate the applicability of the above derived formulas we exploit the trimmed data $Z+$ and $Z-$.

Firstly, we concentrate on the density of the sufficient statistic, MLE estimate and confidence areas, the I-divergence for trimmed data $Z+$. We will consider two independent Pareto distributed variables y_1, y_2 with scale parameter $x_m = 2.03 \times 10^{-6}$. To study aforementioned properties with respect to changes in the unknown parameters γ_i, $i = 1, 2$, we randomly chose two observations $y_1 = 1.45 \times 10^{-5}$ and $y_2 = 4.24 \times 10^{-6}$ within the admissible area: $\gamma_i < -1$, $i = 1, 2$. Results are shown in Fig. 4.5a, 4.5b and 4.5c.



(a) Density of t (b) Density of MLE (c) I-divergence

Figure 4.5: Pareto real data Z+ with saddlepoint densities for sufficient statistic t, MLE estimate and I-divergence.

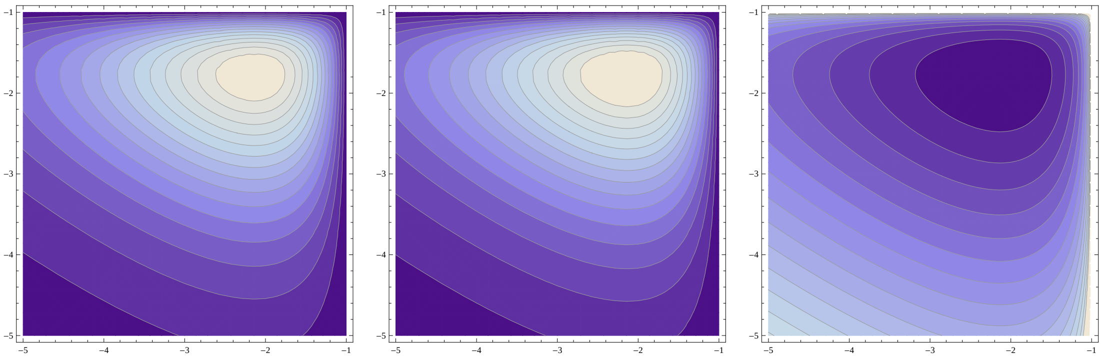
In case of the I-divergence $I_2(\hat{\gamma}_y, \gamma)$, we expect that as a function of two variables it reaches its minimum value, zero, for the arguments:

$$\hat{\gamma}_i = \frac{1}{\log y_i - \log x_m} - 1, \ \ i = 1, 2$$

(see properties of the KL-divergence). Thus, in this case in the point $\hat{\gamma}_1 = -1.51$ and $\hat{\gamma}_2 = -2.36$, which can be seen in 4.5c also brings the view on the changes of the I-divergence with respect to γ_1, γ_2. We can see that plotted I-divergence reaches minimum value for values $\hat{\gamma}_1$ and $\hat{\gamma}_2$. From the principle of maximum likelihood, we expect that the approximated density reaches its maximum values for these values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ (Fig. 4.5a, 4.5b, 4.5c).

The similar analysis was carried out for trimmed data set $Z-$ with values $x_m = 2 \times 10^{-6}$, $y_1 = 4.85 \times 10^{-6}$ and $y_2 = 7.38 \times 10^{-6}$. The results are shown in Fig. 4.6a, 4.6b, 4.6c and corresponding MLE are $\hat{\gamma}_1 = -2.13$, $\hat{\gamma}_2 = -1.77$.

Once we have derived the approximation for the density of $\hat{\gamma}$, it enables us to do more statistical inference on this parameter. One of the biggest advantages of saddlepoint approximation based on I-divergence is that it provides analytical formula for approximation of the density of MLE. Therefore there is no need for any other numerical procedures and the density may be used straighforwardly. In this case not only the density is tractable in explicit form, but also distribution function possesses this desirable property.



(a) Density of t (b) Density of MLE (c) I-divergence

Figure 4.6: Pareto real data Z- with saddlepoint densities for sufficient statistic t, MLE estimate and I-divergence.

4.5 Comparison to the general saddlepoint approximations

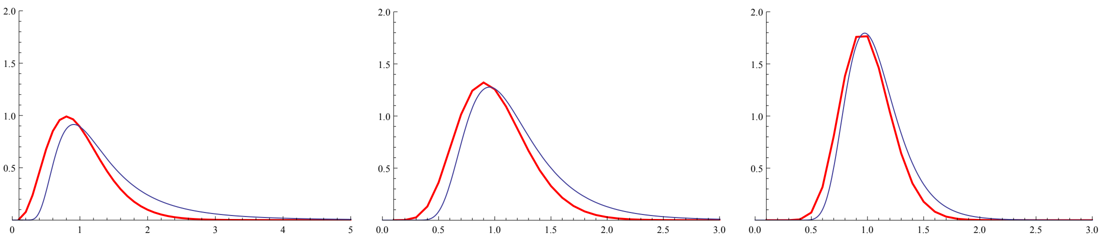
Once we have derived the saddlepoint approximation for the density of MLE in exponential family, it is of interest to compare these results to the original saddlepoint approximations developed in Field and Ronchetti (1990). In order to do so, it is necessary to consider a slightly different setup. Let $Y_1, \ldots, Y_n$ be i.i.d with distribution F_θ and corresponding pdf f_θ. MLE is an M-estimator with

$$\psi(y, \theta) = -\frac{f'(y, \theta)}{f(y, \theta)}.$$

In order to make comparison to our setup we need to restrict ourselves to the situation $\gamma_1 = \gamma_2 = \ldots = \gamma_n$.

From the computational point of view, the biggest difference between original saddlepoint approximations and those based on I-divergence is in the existence of approximation for the density in a closed form. When using saddlepoint approximations for M-estimators, it is often not possible to derive explicit formula of the approximation for the density. It is viable only in special cases already mentioned in the previous chapters, for example for normally distributed random variable. In this case, both methods yield the exact density of MLE. The closed form of the approximation based on I-divergence enables us to compute integrals of the density more precisely. In this case, saddlepoint approximations developed in Field and Ronchetti (1990) yield just a set of values of density in specified points.

Therefore we are not usually able to compare formulae using these two methods and we will present the results of the numerical comparison for different sample sizes. Let us now consider a random sample of size $n = 5, 10$ and 20 from exponential distribution with $\gamma = 1$. In the following figures, for these given sample sizes the saddlepoint approximations based on I-divergence are plotted in blue, approximations based on Field and Ronchetti (1990) are plotted by red line. These two approximations differ, but for growing sample size they begin to coincide.



(a) n = 5 (b) n = 10 (c) n = 20

Figure 4.7: Comparison of saddlepoint approximations for MLE in exponential distribution.

Conclusion

Although saddlepoint methods were introduced into statistic 60 years ago, they still prove to be useful in many fields of statistical inference. The focus point of this thesis was the derivation of the distribution of estimators and testing hypotheses based on saddlepoint methods in quantile regression. This approach yields not only precise approximations to the density of the estimators, but also tests with excellent accuracy in small sample sizes as well as good robustness properties. We derived test statistics for parametric and nonparametric setup (unspecified error distribution in the linear regression model) and compared in a numerical study their performacne to other available alternatives. Unlike vast majority of results based on saddlepoint techniques, the test statistics and approximations of the density introduced in this work are given by explicit formulae, and hence easy to implement.

Beside procedures based on the saddlepoint methods, a test on the value of a regression quantile based on the asymptotic distribution of averaged regression quantiles was proposed. Although this test statistic requires estimation of quantile density function, in a numerical study it proved to be much more precise than classical parametric tests based on likelihood. Also approach to saddlepoint approximations to the density of MLE and sufficient statistic based on information theory was studied. I-divergence as a special case of Kullback-Leibler divergence can be decomposed into a sum of two independent tests based on likelihood ratio, and provides a convenient way of testing the hypothesis on the value of a parameter in exponential family. The results based on I-divergence were applied to real data on methane ebullition.

Bibliography

Barndorf-Nielsen, O. (1978), *Information and Exponential Families In Statistical Theory*, John Wiley & Sons, Ltd., New York.

Bartlett, M. and Kendall, D. (1946), 'The statistical analysis of variance-heterogeneity and the logarithmic transformation', *Supplement to the Journal of the Royal Statistical Society* **8**, 128–138.

Bickel, P. (1974), 'Edgeworth expansions in nonparametric statistics', *The Annals of Statistics* pp. 828–838.

Butler, R. W. (2007), *Saddlepoint Approximations with Applications*, Cambridge University Press, Cambridge, UK.

Daniels, H. (1954), 'Saddlepoint approximations on statistics', *Ann. Math. Statist.* **25**, 631–650.

Dodge, Y. and Jurečková, J. (2000), *Adaptive Regression*, Springer, New York.

Fan, R. and Field, C. (1995), 'Approximations for marginal densities of m-estimators', *The Canadian Journal of Statistics* **23**, 185–197.

Field, C. A. and Hampel, F. R. (1982), 'Small-sample asymptotic distribution of m-estimators of location', *Biometrika* **69**, 29–46.

Field, C. A. and Ronchetti, E. (1990), *Small Sample Asymptotics*, Institute of Mathematical Statistics - Monograph Series, Hayward, CA.

Gutenbrunner, C. and Jurečková, J. (1992), 'Regression rank scores and regression quantiles', *The Annals of Statistics* **20**, 305–330.

Gutenbrunner, C., Jurečková, J., Koenker, R. and Portnoy, S. (1993), 'Tests of linear hypotheses based on regression rank scores', *Journal of Nonparametric Statistics* **2**, 307–331.

Hampel, F. (1974), 'The influence curve and its role in robust estimation', *Journal of the American Statistical Association* **69**, 383–393.

Hampel, F. R. (1973), Some small-sample asymptotics, *in* J. Hájek, ed., 'Proc. Prague Symposium on Asymptotic Statistics', Charles University in Prague, pp. 109–126.

Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, Wiley, NY.

Huber, P. and Ronchetti, E. (2009), *Robust statistics*, Wiley.

Huzurbazar, S. (1999), 'Saddlepoint approximations on statistics', *The American Statistician* **53**, 225–232.

Jaynes, T. (1982), 'On the rationale of maximum entropy methods', *Proc. IEEE* **70**, 939–952.

Jensen, J. L. (1995), *Saddlepoint Approximations*, Oxford University Press, Oxford, UK.

Jordanova, P., Dušek, J. and Stehlík, M. (2013a), 'Microergodicity effects on ebullition of methane modelled by mixed 2 poisson process with pareto mixing variable', *Chemometrics and Intelligent Laboratory Systems* **128**, 124–134.

Jordanova, P., Dušek, J. and Stehlík, M. (2013b), 'Modeling methane emission by the infinite moving average process', *Chemometrics and Intelligent Laboratory Systems* **122**, 40–49.

Jurečková, J. (1999), 'Equivariant estimators and their asymptotic representations', *Tatra Mountains Mathematical Publications* **17**, 1–9.

Jurečková, J. (2010), 'Finite sample distribution of regression quantiles', *Statistics & Probability Letters* **80**, 1940–1946.

Jurečková, J. and Sabolová, R. (2011), 'Finite-sample density and its small sample asymptotic approximation', *Statistics & Probability Letters* **81**, 1311–1318.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**, 33–50.

Koenker, R. W. (1994), Confidence intervals for regression quantiles, *in* M. Hušková and P. Mandl, eds, 'Asymptotic Statistics', Springer-Verlag, New York, pp. 349–359.

Kullback, S. and Leibler, R. A. (1951), 'On information and sufficiency', *Ann. Math. Statist.* **22**, 79–86.

Le Mer, J. and Roger, P. (2001), 'Production, oxidation, emission and consumption of methane by soils: A review', *Eur. J. Soil Biol* **37**, 25–50.

Ma, Y. and Ronchetti, E. (2011), 'Saddlepoint test in measurement error models', *Journal of the American Statistical Association* **106**, 147–156.

Pázman, A. (1993), *Nonlinear statistical Models*, Kluwer Acad. Publ., Dordrecht.

Ronchetti, E., Robinson, J. and Young, G. (2003), 'Saddlepoint approximations and tests based on multivariate m-estimates', *The Annals of Statistics* **31**, 1154–1169.

Ronchetti, E. and Sabolová (2014), Saddlepoint tests for quantile regression. unpublished manuscript.

Ronchetti, E. and Welsh, A. (1994), 'Empirical saddlepoint approximations for multivariate m-estimators', *Journal of the Royal Statistical Society* **56**, 313–326.

Ruppert, D. and Carroll, R. J. (1980), 'Trimmed least squares estimation in the linear model', *J. Amer. Statist. Assoc* pp. 828–838.

Sabolová, R., Sečkárová, V., Dušek, J. and Stehlík, M. (2014), Entropy based statistical inference for methane emissions released from wetland. submitted to Chemometrics and Intelligent Laboratory Systems.

Shannon, C. E. (1948), 'A mathematical theory of communication', *The Bell System Technical Journal* **27**, 379–423.

Shore, J. E. and Johnson, R. (1980), 'Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy', *IEEE Trans. Inf. Theory* **26**, 26–37.

Shukla, P., Pandey, K. and Mishra, V. (2013), 'Environmental determinants of soil methane oxidation and methanotrophs', *Critical Reviews in Environmental Science and Technology* **43**, 1945–2011.

Stehlík, Economou, P., Kiseľák, J. and Richter, W. (2014), 'Kullback–leibler life time testing', *Applied Mathematics and Computation* **240**, 122 – 139.

Stehlík, M. (2003), 'Distributions of exact tests in the exponential family', *Metrika* **57**, 145–164.

Stehlík, M., Střelec, L. and Thulin, M. (2013), 'On robust testing for normality in chemometrics', *Chemometrics and Intelligent Laboratory Systems* **130**, 98–108.

Yari, G. and Borzadaran, G. (2010), 'Entropy for pareto-types and its order statistics distributions', *Commun. Inf. Syst.* **10**, 193–202.

List of Tables

List of Figures

List of Abbreviations

CLT	central limit theorem
KL-divergence	Kullback-Leibler divergence
log	natural logarithm
LR	likelihood ratio
LSE	least squares estimator
MLE	maximum likelihood estimator
pdf	probability density function
pmf	probability mass function