Technische Universität Ilmenau
Institut für Mathematik

Preprint No. M **17/13**

Confidence sets in decision problems with kernel density estimators

Silvia Vogel, Stefan Seeger

2017

URN: urn:nbn:de:gbv:ilm1-2017200528

Impressum:
Hrsg.: Leiter des Instituts für Mathematik
Weimarer Straße 25
98693 Ilmenau
Tel.: +49 3677 69-3621
Fax: +49 3677 69-3270
http://www.tu-ilmenau.de/math/

Confidence Sets in Decision Problems with Kernel Density Estimators

Silvia Vogel, Stefan Seeger

Keywords: kernel density estimator, uniform confidence band, mode estimator, high density region, chance constraints

Mathematics Subject Classification: 62G07, 62G15, 90C15, 90C31

Abstract

Random approximations for a deterministic optimization problem occur in many situations. Unknown parameters or probability distributions in real-life decision problems are replaced with estimates; more and more solution algorithms use random steps. Moreover, many estimation procedures in statistics are random optimization problems, which can be supplemented with a deterministic limit problem.

Confidence sets for solution sets and level sets of a deterministic decision problems can be derived on the base of suitable uniform concentration-of-measure results for sequences of random functions. In the present paper approximations with kernel density estimators are considered and concentration-of-measure results for these estimators are provided. The results are employed to derive confidence sets for modes and high density regions of probability distributions. Furthermore, applications in stochastic programming are investigated. It is shown how the assertions can be utilized to derive uniform concentration-of-measure results for chance constraints and functions which are expectations.

1 Introduction

Usually, real-life decision problems are optimization problems which are fraught with uncertainties. Decision makers estimate the unknown parameters or probability distributions and solve the problems as they then arise. This is particularly true for stochastic programming problems, which naturally heavily depend on the underlying probability distribution. Unfortunately, sometimes small deviations in the objective functions and/or the constraints of a decision problem can result in relatively large deviations in the solutions. Hence there is a need for methods that take into account that the data are uncertain and so are the results.

There are many approaches that try to cope with this challenge. Two major techniques were developed by Jitka Dupačová and are successfully applied to the present day, the minimax approach ([33], [24]) and the contamination technique ([4], [5]).

An approach which is closely related to the minimax approach takes into account the 'ambiguity' of the probability measure and tries to find a decision which is 'optimal' with respect to a suitably chosen set of probability measures, see e.g. [18], [10].

Useful information is also given by convergence conditions as considered in [6]. Usually one can directly use convergence results for the estimates involved and derive convergence properties for solution sets etc. ([23], [17], [28]). However, one often faces the situation, that the generation of new samples is expensive, if not impossible. Then quantitative results for a fixed sample size n are asked for. Quantitative stability results have been provided in terms of distances between the underlying probability measures or the parameters ([20], [19], [18]).

Another approach uses sets which cover the true solutions or optimal values of stochastic programming problems with a high probability. Such 'confidence sets' for the solutions of the decision problems can be derived if confidence regions for the unknown objective functions and constraint sets are available ([17], [29]). Often the functions involved depend on a probability distribution, as they are expectations of random functions or chance constraints.

Approximations of the probability measure by the empirical measure are considered in the framework of estimated probability distributions, but also for numerical reasons, above all in the context of Sample Average Approximation ([23], [14], [2], [32]). For chance constraints, however, it has the disadvantage that discontinuous functions come into play. If the random variables are known to be continuously distributed, one can benefit from the smoothing effect of the density estimators, particularly for numerical reasons.

In the present paper the focus is on approximations of the probability measure with kernel density estimators. Density estimation with kernels has a long tradition in statistics and is widely spread. Kernel estimates are well investigated, especially with respect to their convergence properties, cf. [21], [15], [12], [26], [7]. Parametric density estimation or density estimation with Epi-splines ([22]) are further interesting methods, but will not be considered here. However, given concentration-of-measure results for these estimators, one can proceed as in Section 3 of this paper to derive corresponding results.

We are interested in the non-asymptotic point of view, particularly in confidence sets. In parametric statistics confidence sets are usually derived from statistics with a known distribution. If no such distribution is available one usually makes do with the asymptotic distribution. Asymptotic confidence sets have also been considered in stochastic programming ([16], [9], [2]) and in the context of level sets for kernel density estimators ([11]). Here we pursue another approach. Exploiting quantified convergence notions for sequences of random functions and sequences of random sets one can derive so-called universal confidence sets. The method yields for each sample size n a conservative confidence set, i.e. a set which covers the true set with a prescribed probability.

The crucial assumptions in the approach are concentration-of-measure properties for sequences of random functions. It is the aim of the present paper to provide corresponding assertions for approximations on the base of density estimators.

The results can not only be applied to problems arising from random approximations to mathematical, particularly stochastic, programming problems. Already by Jitka Dupačová and Roger J-B Wets [6] and later on in [28] it was pointed out that methods originally developed in the framework of stochastic programming can also contribute to statistical estimation theory. This holds equally for kernel density estimators. The investigation of key properties of the density itself can benefit from the approach investigated in the present paper. Thus it is possible to derive confidence sets for modes or high density regions.

For the general approach investigated in [29] uniform concentration-of-measure properties of the form (1) below are required. Unfortunately, to the best of our knowledge, so far no such result has been proved for kernel density estimators. We therefore provide a result which can be used to determine confidence sets for high density regions and modes. Moreover, the uniform confidence bands can help to judge to what extent 'desirable' densities, e.g. normal distributions with special properties in stochastic programming problems, fit into the confidence band.

Concentration-of-measure results for integrals of densities have been derived in [3] under differentiability conditions for the density. We will extend these assertions and show how they can be exploited for the investigation of functions which are integrals and particular chance constraints. Note that important risk measures can be represented as integrals.

The approach which will be considered is usually easy to apply, once the needed convergence properties have been proved. We provide a short example.

Suppose that we would like to investigate an unknown distribution represented by a density function f_0. Assume that a density estimator f_n with the following property is available:

$$\sup_{n\in N} P(\sup_{z\in\mathbb{R}^m} |f_n(z) - f_0(z)| \geq \beta_{n,\kappa}) \leq \mathcal{H}(\kappa). \qquad (1)$$

Here $\mathcal{H}(\kappa)$ denotes a function with the property $\lim\limits_{\kappa\to\infty} \mathcal{H}(\kappa) = 0$ and $(\beta_{n,\kappa})_{n\in N}$ denote sequences of positive real numbers with $\lim\limits_{n\to\infty} \beta_{n,\kappa} = 0$ for each $\kappa > 0$.

Inequality (1) can immediately be used to derive a universal confidence band for f_0, but, moreover, offers the possibility to obtain confidence sets for level sets, argmax sets etc. Here we consider the derivation of a confidence area for a level set $M^\delta := \{z \in \mathbb{R}^m : f_0(z) \geq \delta\}$, $\delta > 0$, of the density function f_0. To avoid additional technical considerations we assume that f_0 is upper semicontinuous. Further assertions are presented in Section 3.

Consider the random sets $M^\delta_{n,\kappa} := \{z \in \mathbb{R}^m : f_n(z) \geq \delta - \beta_{n,\kappa}\}$, $n \in N$. If, for a fixed $n \in N$, there is a $z_n \in M^\delta$ which does not belong to $M^\delta_{n,\kappa}$, then

$f_0(z_n) \geq \delta$, but $f_n(z_n) < \delta - \beta_{n,\kappa}$. Hence $f_n(z_n) - f_0(z_n) < -\beta_{n,\kappa}$ and by (1) the probability of the event $M^\delta \setminus M^\delta_{n,\kappa} \neq \emptyset$ can be bounded by $\mathcal{H}(\kappa)$. Consequently

$$\sup_{n \in N} P(M^\delta \setminus M^\delta_{n,\kappa} \neq \emptyset) \leq \mathcal{H}(\kappa).$$

In order to derive a confidence set for M^δ with a prescribed level $1-\eta$ one determines κ_0 such that $\mathcal{H}(\kappa_0) \leq \eta$. Then for each sample size n the set M^δ_{n,κ_0} covers the true level set M^δ at least with probability $1-\eta$. Note that no knowledge about the whole distribution or the asymptotic distribution is needed and a confidence set for each sample size n can be derived. In fact, for this application we need only the weaker one-sided assertion $\sup\limits_{n \in N} P(\inf\limits_{z \in \mathbb{R}^m} (f_n(z) - f_0(z)) \leq -\beta_{n,\kappa}) \leq \mathcal{H}(\kappa)$. In order to find a meaningful level δ mode estimators can be employed.

The paper is organized as follows: In Section 2 we introduce the kernel density estimator and provide concentration-of-measure results, especially of the form (1), under different assumptions. In Section 3 we show how confidence sets for modes and high density regions can be derived. Furthermore we consider confidence bands for functions which are integrals with respect to a density and confidence sets for probabilistic constraints. In Section 4 we provide the proofs for Section 2.

2 Concentration-of-Measure for the Kernel Density Estimator

Let $Z_1, Z_2, \ldots$ be i.i.d. random vectors on a probability space $[\Omega, \Sigma, P]$ with values in R^m. The probability distribution P_Z is assumed to have a density f_0. We consider the kernel density estimator of f_0 based on $Z_1, \ldots, Z_n,\ n \geq 1$,

$$f_n(z) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{z - Z_i}{h_n^{1/m}}\right).$$

where K is a kernel and $h_n > 0$ is the bandwidth. Sometimes, in order to make clear which variables are random, we will use the extended writing

$$f_n(z, \omega) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{z - Z_i(\omega)}{h_n^{1/m}}\right).$$

The kernel estimators do not have the most general form. Instead of a uniform bandwidth h_n 'individual' bandwidths or bandwidth matrices can be used. The approach can be extended to this case, see [25]. In order not to overload the presentation with technical details we use uniform bandwidths. Furthermore, for the same reason, here we will not consider corrections for f_n if the support does not equal $\mathbb{R}^m$.

Suitable multivariate kernels K are often obtained as products of univariate kernels $\tilde{K}$: $K(u) = \prod_{i=1}^{m} \tilde{K}(u_i),\ u = (u_1, \ldots, u_m)^T$. Another method uses $K(u) = c_K \tilde{K}(||u||_2)$ with $c_K^{-1} = \int_{\mathbb{R}^m} \tilde{K}(||u||_2)du$ and the Euclidean norm $|| \ldots ||_2$.

It is well known from statistics that in order to ensure consistency of the estimators also assumptions on the function f_0 have to be imposed ([26]). Naturally, we also need conditions of that kind. When employing formulas like (1), it is sometimes useful to consider $\sup_{z \in \mathcal{Z}} \ldots$ for a Borel set $\mathcal{Z} \subset \mathbb{R}^m$ instead of $\sup_{z \in \mathbb{R}^m} \ldots$ Then also the assumptions concerning f_0 can be restricted to the set $\mathcal{Z}$ or a suitable neighborhood $U\mathcal{Z}$. $|| \ldots ||$ denotes a suitable vector norm. We will derive assertions of the form (1) under a differentiability assumption and for Hölder continuous functions f_0.

($Vf2$-$\mathcal{Z}$) f_0 is in $C^2(U\mathcal{Z})$ and its partial derivatives of order 1 and 2 are bounded, especially there exist $C_3 < \infty$ such that
$\forall i, j \in \{1, \ldots, m\} : \sup_{z \in U\mathcal{Z}} \left| \frac{\partial^2 f_0(z)}{\partial z_i \partial z_j} \right| \leq C_3$.

(VfH-$\mathcal{Z}$) There exist $L \geq 0$ and $\alpha \in (0, 1]$ such that
$\forall (t_1, t_2)^T \in U\mathcal{Z} : |f_0(t_1) - f_0(t_2)| \leq L||t_1 - t_2||^{\alpha}$.

One may argue that the true density is not known. However, often suitable bounds are available. Furthermore, as we aim at confidence sets, foregoing confidence sets for the bounds can be incorporated. Note that also results which assume higher order differentiability can be derived. They allow for better convergence rates.

For uniform concentration-of-measure assertions we need the Fourier transform k of the kernel K

$$k(u) := \int_{\mathbb{R}^m} e^{iu^T y} K(y) dy \quad \forall\ u \in \mathbb{R}^m$$

and an assumption concerning k.

The following conditions for the kernel K will be considered. (K5) and (K6) are used only in connection with ($Vf2$-$\mathcal{Z}$) and (K8) in connection with (VfH-$\mathcal{Z}$).

$$
\begin{array}{ll}
(K1) & K(u) = K(-u) \ \forall u \in \mathbb{R}^m, \\
(K2) & \int_{\mathbb{R}^m} K(u)du = 1, \\
(K3) & \sup_{z\in\mathbb{R}^m} K(z) - \inf_{z\in\mathbb{R}^m} K(z) =: C_1 < \infty, \\
(K4) & \int_{\mathbb{R}^m} |k(u)|\, du =: C_2 < \infty, \\
(K5) & \int\limits_{\mathbb{R}^m} u_i K(u)du = 0 \ \ \forall i = 1,\dots,m, \ u = (u_1,\dots,u_m)^T, \\
(K6) & \int_{\mathbb{R}^m} u^T u K(u)du =: C_4 < \infty, \\
(K7) & \int_{\mathbb{R}^m} |K(u)|\, du =: C_5 < \infty, \\
(K8) & \int_{\mathbb{R}^m} ||u||^\alpha |K(u)| du =: \mu_\alpha < \infty \text{ where } \alpha \in (0,1], \\
(K9) & \int_{\mathbb{R}^m} (K(u))^2 du =: C_6 < \infty.
\end{array}
$$

(K2) and (K3) imply that $\mathbb{E}f_n(z) = \int\limits_{\omega\in\Omega} f_n(z,\omega)dP(\omega) = \int\limits_{\mathbb{R}^m} K(z-y)f_0(y)dy$ exists.

Firstly, we present a uniform concentration-of-measure assertion. It is the base for the derivation of confidence sets for high density regions and modes. First investigations of the uniform case in the univariate setting can be found in [8].

Theorem 1 (Uniform Concentration-of-Measure)
Let the conditions (K1) - (K4) be satisfied.

(i) If additionally the assumptions (Vf2-$\mathcal{Z}$), (K5), and (K6) are fulfilled, the inequality

$$\sup_{n\in N} P(\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| \geq \tfrac{\kappa}{\sqrt{n}h_n} + \tfrac{C_2}{(2\pi)^m\sqrt{n}h_n} + \tfrac{1}{2}C_3C_4h_n^{2/m}) \leq 2e^{\frac{-2\kappa^2}{C_1{}^2}}$$

holds.

(ii) If additionally the assumptions (VfH-$\mathcal{Z}$) and (K8) are fulfilled, the inequality

$$\sup_{n\in N} P(\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| \geq \tfrac{\kappa}{\sqrt{n}h_n} + \tfrac{C_2}{(2\pi)^m\sqrt{n}h_n} + L\mu^\alpha h_n^{\alpha/m}) \leq 2e^{\frac{-2\kappa^2}{C_1{}^2}}$$

holds.

Consequently, in order to derive useful sequences $(\beta_{n,\kappa})_{n\in N}$, $\kappa > 0$, the assumptions

(B1) $\lim_{n\to\infty} h_n = 0$ and

(B2) $\lim_{n\to\infty} nh_n^2 = \infty$

have to be imposed.

Sometimes also a 'pointwise' version of Theorem 1 can be useful. Then we do not need the second summand in the formulas for $\beta_{n\kappa}$, see Section 4.

Theorem 2 (Pointwise Concentration-of-Measure)
Let the conditions (K1), (K2), and (K3) be satisfied.

(i) If additionally the assumptions (Vf2-$\{z\}$), (K5), and (K6) are fulfilled, the inequality
$\sup_{n\in N} P(|f_n(z) - f_0(z)| \geq \frac{\kappa}{\sqrt{nh_n}} + \frac{1}{2}C_3C_4h_n^{2/m}) \leq 2e^{\frac{-2\kappa^2}{C_1{}^2}}$
holds.

(ii) If additionally the assumptions (VfH-$\{z\}$) and (K8) are fulfilled, the inequality $\sup_{n\in N} P(|f_n(z) - f_0(z)| \geq \frac{\kappa}{\sqrt{nh_n}} + L\mu^\alpha h_n^{\alpha/m}) \leq 2e^{\frac{-2\kappa^2}{C_1{}^2}}$
holds.

Now integrals of densities over a bounded Borel set $\mathcal{T}$ will be considered. For $M \subset \mathbb{R}^m$ the set U_hM is defined by $U_hM := \{y \in \mathbb{R}^m : \inf_{x\in M} ||x-y|| < h\}$. $\bar{U}_hM$ means its closure. We will assume that $U_{h_n}\mathcal{Z} \subset U\mathcal{Z}$ and $U_{h_n}\mathcal{T} \subset U\mathcal{T}$ whenever these conditions are needed. Otherwise the assertions have to be adjusted, which is usually possible. λ denotes the Lebesgue measure.

Theorem 3 (Concentration-of-Measure of Integrals)
Let the conditions (K1), (K2), (K7), and (K9) be satisfied.

(i) If additionally the assumptions (Vf2-$\mathcal{T}$) and (K6) are fulfilled, the inequality

$$\sup_{n\in N} P(\int_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz \geq \beta_{n,\kappa}) \leq e^{\frac{-\kappa}{2C_5^2}} \qquad (2)$$

with $\beta_{n,\kappa} = \frac{\kappa}{\sqrt{n}} + \frac{1}{\sqrt{nh_n}}\sqrt{\lambda(U_{h_n}\mathcal{T})C_6} + \frac{1}{2}C_3C_4h_n^{2/m}$
holds.

(ii) If additionally the assumptions (VfH-$\mathcal{T}$) and (K8) are fulfilled, the inequality (2) holds with
$\beta_{n,\kappa} = \frac{\kappa}{\sqrt{n}} + \frac{1}{\sqrt{nh_n}}\sqrt{\lambda(U_{h_n}\mathcal{T})C_6} + L\mu^\alpha h_n^{\alpha/m}\lambda(\mathcal{T})$.

3 Applications

3.1 Confidence Sets for High Density Regions and Modes

Given a uniform concentration-of-measure result of the form (1), a high density region can be derived immediately as shown in the introduction. However, the obtained approximation may have a complicated shape. Further approximations by a superset of a simpler form, e.g. via a parametric function, can be helpful ([25]).

Moreover, there is also another possibility, which uses a quantified one-sided version of Kuratowski-Painlevé-convergence in probability of sequences of random sets. Abbreviating we will speak of the KP-approach. Because a level set can be described by inequality constraints, the results about the approximation of constraint sets in [29] can be employed. Generally speaking, the approach works as follows. One determines the set under consideration for the approximate problem and adds a suitable ball to each point. Thus it remains to find a formula for the radius of the balls. The determination of the radius requires some knowledge about the true problem. For instance, when considering high density regions, a bound for the growth of the density near the boundary of the high density region is utilized.

Firstly, we quote a result from [29] in a specialized form. Let $g_0|\mathbb{R}^m \to \mathbb{R}^1$ be a function which is lower semicontinuous in all points $z \in \mathbb{R}^m$, and define

$$\Gamma_0 := \{z : \; g_0(z) \leq 0\}.$$

Γ_0 is assumed to be non-empty. The function g_0 is approximated by a sequence $(g_n)_{n\in N}$ of functions $g_n|\mathbb{R}^m \otimes \Omega \to \mathbb{R}^1$ which are $(\mathcal{B}^m \otimes \Sigma, \mathcal{B}^1)$-measurable. $\mathcal{B}^r$ denotes the σ-field of Borel sets of $\mathbb{R}^r$. Furthermore, we assume that the functions $g_n(\cdot, \omega)$ are lower semicontinuous for almost all $\omega \in \Omega$.

Eventually, the approximate constraint set Γ_n is defined by

$$\Gamma_n(\omega) := \{z \in \mathbb{R}^m : \; g_n(z, \omega) \leq 0\}.$$

Under our assumptions Γ_n is a closed-valued measurable multifunction.

The crucial condition in the next theorem is condition (CO1). Here the results of the foregoing section can be exploited. Condition (CO2) is a quantified 'inner point condition'. Let, for a given $\varepsilon > 0$, $CI(\varepsilon) := \Gamma_0 \setminus U_\varepsilon(\mathbb{R}^m \setminus \Gamma_0)$. The requirement $\Gamma_0 \subset \bar{U}_\varepsilon CI(\varepsilon)$ is needed since we allow for rather general sets Γ_0. For convex sets the condition can be considerably simplified.

Furthermore, we use the following denotations. B is the set of sequences of positive numbers that converge monotonously to zero. H denotes the set of functions $\mathcal{H}|\mathbb{R}^1_+ \to \mathbb{R}^1_+$ with $\lim\limits_{\kappa\to\infty} \mathcal{H}(\kappa) = 0$. Λ is the set of functions $\tilde{\lambda}|\mathbb{R}^1 \to \mathbb{R}^1$ which are right-continuous, non-decreasing, non-constant, and have the property $\tilde{\lambda}(0) = 0$. By the superscript $^{-1}$ we denote their inverses: $\tilde{\lambda}^{-1}(y) := \inf\{x \in \mathbb{R}^1 :$

$\tilde{\lambda}(x) \geq y\}$.

Proposition. Assume that the following conditions are satisfied:

(CO1) There exist a function $\mathcal{H} \in H$ and for all $\kappa > 0$ a sequence $(\beta_{n,\kappa})_{n\in N} \in B$ such that
$\sup_{n\in N} P\{\omega : \sup_{z\in\Gamma_0}(g_n(z,\omega) - g_0(z)) \geq \beta_{n,\kappa}\} \leq \mathcal{H}(\kappa)$.

(CO2) There exist an $\tilde{\varepsilon} > 0$ and a function $\mu_c \in \Lambda$ such that for all $0 < \varepsilon \leq \tilde{\varepsilon}$
$\Gamma_0 \subset \bar{U}_\varepsilon CI(\varepsilon)$ and $\forall z \in CI(\varepsilon): \; g_0(z) \leq -\mu_c(\varepsilon)$.

Then for all $\kappa > 0$ and $\beta^{(oc)}_{n,\kappa} = \mu_c^{-1}(2\beta_{n,\kappa})$ the relation
$\sup_{n\in N} P\{\omega : \; \beta^{(oc)}_{n,\kappa} \leq 2\tilde{\varepsilon}$ and $\Gamma_0 \setminus U_{\beta^{(oc)}_{n,\kappa}}\Gamma_n(\omega) \neq \emptyset\} \leq \mathcal{H}(\kappa)$
holds.

The condition $\beta^{(oc)}_{n,\kappa} \leq 2\tilde{\varepsilon}$ is no restriction. It is only used to indicate that arbitrary large values for $\beta^{(oc)}_{n,\kappa}$ do not make sense.

The lemma is an example of a so-called outer approximation, which yields confidence sets with a confidence level $1-\eta$ as follows. Let κ_0 such that $\mathcal{H}(\kappa_0) \leq \eta$. Then, for each $n \in N$, $U_{\beta^{(oc)}_{n,\kappa_0}}\Gamma_n(\omega)$ is a confidence set for Γ_0 with confidence level $1-\eta$.

In order to judge the quality of these approximations, inner approximations ([29]) or subset approximations could be employed. However, inner approximations need not provide subsets. Fortunately, often similarly to the example in the introduction subset approximations can be derived.

High density regions are upper level sets. Hence we are interested in sets

$$\tilde{\Gamma}_0 := \{z \in \mathbb{R}^m : f_0(z) \geq \delta\}$$

for a suitable $\delta > 0$. Furthermore, let

$$\tilde{\Gamma}_n := \{z \in \mathbb{R}^m : f_n(z) \geq \delta\}.$$

Theorem 4 below provides an example of an assertion which may be derived combining the above proposition with Theorem 1. Many important densities are unimodal, i.e. they have one local maximum point only. Univariate unimodal densities are quasiconcave, consequently the upper level sets are convex and $\tilde{\Gamma}_0 \subset \bar{U}_\varepsilon CI(\varepsilon)$ is satisfied. We provide a result which exploits part (i) of Theorem 1. Part (ii) yields a corresponding result.

Theorem 4 (High Density Regions - KP-Approach)
Let the assumptions of Theorem 1(i) be satisfied and assume that f_0 is upper semi-continuous and quasiconcave. Furthermore, suppose that the following condition is fulfilled:

(CO2-u) *There exist an* $\tilde{\varepsilon} > 0$ *and a function* $\mu_c \in \Lambda$ *such that for all* $0 < \varepsilon \leq \tilde{\varepsilon}$ $\forall z \in CI(\varepsilon): \ f_0(z) \geq \delta + \mu_c(\varepsilon)$.

Then for all $\kappa > 0$ *and* $\beta^{(oc)}_{n,\kappa} = \mu_c^{-1}(2\frac{\kappa}{\sqrt{n}h_n} + 2\frac{C_2}{(2\pi)^m\sqrt{n}h_n} + C_3C_4h_n^{2/m})$ *the relation* $\sup_{n\in N} P\{\omega : \ \beta^{(oc)}_{n,\kappa} \leq 2\tilde{\varepsilon}$ *and* $\tilde{\Gamma}_0 \setminus U_{\beta^{(oc)}_{n,\kappa}}\tilde{\Gamma}_n(\omega) \neq \emptyset\} \leq 2e^{\frac{-\kappa^2}{2C_1{}^2}}$ *holds.*

Estimates for modes, i.e. local maxima of a density, have been considered for a long time, see e.g. [31] and the references quoted there. As mentioned, often modes are unique which means that we have a single-valued solution of a maximization problem. Sometimes single-valuedness is imposed via an 'identifiability' condition.Thus we can use a result about a so-called inner Kuratowski-Painlevé-approximation of solution sets. Inner approximations, in general, can be derived under weaker conditions than outer approximations. If the sets under consideration are single-valued, inner approximations are also outer approximations. In terms of the solution sets
$\Psi_n(\omega) := \{z \in R^m : f_n(z,\omega) \geq f_n(\tilde{z},\omega) \ \forall \tilde{z} \in \mathbb{R}^m\}, \ n \in N,$ and
$\Psi_0 := \{z \in R^m : f_0(z) \geq f_0(\tilde{z}) \ \forall \tilde{z} \in \mathbb{R}^m\}$
this means that

$$\sup_{n\in N} P\{\omega : \ \Psi_n(\omega) \setminus U_{\beta_{n,\kappa}}\Psi_0 \neq \emptyset\} \leq \mathcal{H}(\kappa)$$

implies

$$\sup_{n\in N} P\{\omega : \ \Psi_0 \setminus U_{\beta_{n,\kappa}}\Psi_n(\omega) \neq \emptyset\} \leq \mathcal{H}(\kappa).$$

Usually mode estimation problems do not take constraints into account. Therefore we will provide two assertions for the solution of an unconstrained optimization problem. One result, which belongs to the context of Kuratowski-Painlevé-approximations, needs a growth condition for the objective function. The relaxation version, which is similar to the example in the introduction, can cope without any knowledge about the true function. The first assertion could be derived from the more general results for minimization problems in [29]. Because we do not consider constraints, the optimization problem is much easier and we provide the simple direct proof. Note that modes which are not single-valued or maxima of constrained optimization of a density can be treated using the results of [29]. Sufficient conditions for assumption (Vf) below are provided by Theorem 1.

Theorem 5 (Mode - KP-Approach)
Assume that the following assumptions are satisfied:

(Vf) *There exist a function* $\mathcal{H} \in H$ *and to all* $\kappa > 0$ *a sequence* $(\beta_{n,\kappa})_{n\in N} \in B$ *such that* $\sup_{n\in N} P\{\omega : \ \sup_{z\in\mathbb{R}^m} |f_n(z,\omega) - f_0(z)| \geq \beta_{n,\kappa}\} \leq \mathcal{H}(\kappa)$.

(G) There exists a function $\mu \in \Lambda$ *such that for all* $\varepsilon > 0$
$\forall\, z \notin U_\varepsilon \Psi_0 :\ f_0(z) \le \Phi_0 - \mu(\varepsilon)$.

Then for all $\kappa > 0$ *and* $\beta_{n,\kappa}^{(m)} = \mu^{-1}(2\beta_{n,\kappa})$ *the relation*
$\sup_{n \in N} P\{\omega :\ \Psi_n(\omega) \setminus U_{\beta_{n,\kappa}^{(m)}} \Psi_0 \neq \emptyset\} \le \mathcal{H}(\kappa)$
holds. If $\Psi_n,\ n \in N$, *and* Ψ_0 *are single-valued, also*
$\sup_{n \in N} P\{\omega :\ \Psi_0 \setminus U_{\beta_{n,\kappa}^{(m)}} \Psi_n(\omega) \neq \emptyset\} \le \mathcal{H}(\kappa)$
is fulfilled.

Proof of Theorem 5. Assume that there are an $n \in N$, a $\kappa > 0$ and an $\omega \in \Omega$ such that $\Psi_n(\omega) \setminus U_{\beta_{n,\kappa}^{(m)}} \Psi_0 \neq \emptyset$. Hence there is a $z_n(\omega) \in \Psi_n(\omega)$ which does not belong to $U_{\beta_{n,\kappa}^{(m)}} \Psi_0$. According to (G), for $z_0 \in \Psi_0$ we obtain $f_0(z_n) \le \Phi_0 - \mu(\beta_{n,\kappa}^{(m)}) \le \Phi_0 - 2\beta_{n,\kappa} = f_0(z_0) - 2\beta_{n,\kappa}$. Furthermore, we have $f_n(z_0) \le f_n(z_n)$. Consequently $f_0(z_n) - f_n(z_n) \le f_0(z_0) - f_n(z_0) - 2\beta_{n,\kappa}$ and either $f_n(z_n) \ge f_0(z_n) + \beta_{n,\kappa}$ or $f_n(z_0) \le f_0(z_0) - \beta_{n,\kappa}$. In both cases ω belongs to the set $\{\tilde{\omega} :\ \sup_{z \in \mathbb{R}^m} |f_n(z, \tilde{\omega}) - f_0(z)| \ge \beta_{n,\kappa}\}$ and (Vf) can be exploited. □

Now we turn to the relaxation approach. Let Φ_0 denote the maximal value of the true density and Φ_n the maximal value of the density estimator. The following theorem can be used to derive a confidence set for the mode ($a = 1$) or a confidence set for a high density region. In comparison to the example in the introduction, the level is not fixed, but related to the value of the mode.

Theorem 6 (Relaxation)
Assume that condition (Vf) is satisfied. Then for all $\kappa > 0,\ a \in (0, 1]$,
$\Psi_{n,\kappa}^{R,a}(\omega) := \{z \in \mathbb{R}^m : f_n(z, \omega) \ge a\Phi_n(\omega) - (1 + a)\beta_{n,\kappa}\}$, *and*
$\Psi_0^a := \{z \in \mathbb{R}^m : f_0(z) \ge a\Phi_0\}$ *the relation*
$\sup_{n \in N} P\{\omega :\ \Psi_0^a \setminus \Psi_{n,\kappa}^{R,a}(\omega) \neq \emptyset\} \le 2\mathcal{H}(\kappa)$
holds.

Proof. Assume that there are an $n \in N$, a $\kappa > 0$ and an $\omega \in \Omega$ such that $\Psi_0^a \setminus \Psi_{n,\kappa}^{R,a}(\omega) \neq \emptyset$. Hence there is a $z_0(\omega) \in \Psi_0^a$ which does not belong to $\Psi_{n,\kappa}^{R,a}(\omega)$. This implies $f_n(z_0(\omega)) < a\Phi_n(\omega) - (1 + a)\beta_{n,\kappa}$. Let $z_n(\omega)$ be such that $f_n(z_n(\omega), \omega) = \Phi_n(\omega)$. Consequently, because of $f_0(z_0) \ge a\Phi_0 \ge af_0(z_n)$ and $f_n(z_0) < af_n(z_n) - (1 + a)\beta_{n,\kappa}$, we have $f_n(z_0) - f_0(z_0) < af_n(z_n) - af_0(z_n) - (1 + a)\beta_{n,\kappa}$ and we can conclude as in the proof to the foregoing theorem. □

3.2 Functions which are Expectations with Respect to a Density

Objective functions in stochastic programming problems are often expectations of random functions. Important risk measures like Average Value-at-Risk fit into this framework.

In this section we consider decision problems with a decision variable $x \in \mathbb{R}^p$ and objective or constraint functions of the form

$$h_0(x) = \int\limits_{\mathbb{R}^m} \varphi(x,z) f_0(z) dz = \mathbb{E}\varphi(x, Z)$$

where Z is a random variable with values in $\mathbb{R}^m$ and probability distribution P_Z, which has a density f_0. φ is supposed to be Borel measurable and continuous in each x for almost all z. The discontinuity set $D(x) = \{z : \ \varphi(\cdot, z)$ is discontinuous in $x\}$ can vary with x.

Furthermore, we assume that to each $x \in \mathbb{R}^p$ there is a neighborhood $U_{\hat{\varepsilon}}\{x\}$ such that $\mathbb{E} \sup\limits_{\tilde{x} \in U_{\hat{\varepsilon}}\{x\}} |\varphi(\tilde{x}, Z)|$ exists. $\hat{\varepsilon}$ can depend on x.

We assume that the function φ is known and the density is approximated by the considered density kernel estimator f_n. Hence we have

$$h_n(x) = \int\limits_{\mathbb{R}^m} \varphi(x,z) f_n(z) dz.$$

We aim at conditions of the form

$$\sup_{n \in N} P(\sup_{x \in M} |h_n(x) - h_0(x)| \geq \beta_{n,\kappa}) \leq \mathcal{H}(\kappa)$$

where M denotes a Borel subset of $\mathbb{R}^p$. As the functions φ and f_0 usually have different properties for different subsets of the support of f_0, it can be useful to consider a partition $\{A_1, \ldots, A_k\}$ of $\mathbb{R}^m$. This partition may be refined with increasing n, similar to the 'pointwise approach' in [30]. The following inequality can be used:

$$\begin{aligned}\sup_{x \in M} |h_n(x) - h_0(x)| &= \sup_{x \in M} |\sum_{i=1}^{k} \int\limits_{A_i} \varphi(x,z)(f_n(z) - f_0(z)) dz| \\ &\leq \sup_{x \in M} \sum_{i=1}^{k} [\sup_{z \in A_i} |\varphi(x,z)| \sum_{i=1}^{k} \int\limits_{A_i} |f_n(z) - f_0(z)| dz].\end{aligned}$$

Hence the results of Theorem 3 for $\int\limits_{A_i} |f_n(z) - f_0(z)| dz$ can be utilized. Because the Lebesgue measure comes into play, usually a bounded support of the density is needed. In real-life situations this assumption is often satisfied.

Now we consider probabilistic constraints. Let J be a finite index set and

$$\hat{\Gamma}_0 := \{x \in \mathbb{R}^p : \ P_Z\{z : \hat{\gamma}^j(x,z) \leq 0\} \geq \eta^j, \ j \in J\}, \ 0 < \eta^j < 1.$$

$\hat{\gamma}^j | \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^1, \ j \in J$, are measurable functions such that $\hat{\gamma}^j(\cdot, z)$ is lower semicontinuous for P_Z-almost all z. For sake of simplicity we confine to individual chance constraints. Joint chance constraints can be treated in a similar way. In order to exploit the proposition we can use $g_n(x) = \max_{j\in J} \hat{g}_n^j(x),\ n \in N \cup \{0\}$.

With the **1**-function (or indicator function), the set $M^j(x) = \{z \in \mathbb{R}^m : \hat{\gamma}^j(x,z) \leq 0\}$, and

$$\hat{g}_0^j(x) = \eta^j - \mathbb{E}\mathbf{1}_{M^j(x)}(Z) = \eta^j - \int\limits_{M^j(x)} f_0(z)dz$$

the set $\hat{\Gamma}_0$ can also be written in the form $\hat{\Gamma}_0 = \{x \in \mathbb{R}^p :\ \hat{g}_0^j(x) \leq 0,\ j \in J\}$.

Approximating P_Z with a density estimator f_n we obtain

$$\hat{g}_n^j(x,\omega) = \eta^j - \int\limits_{\mathbb{R}^m} \mathbf{1}_{M^j(x)}(z) f_n(z)dz = \eta^j - \int\limits_{M^j(x)} f_n(z)dz.$$

We could employ the foregoing considerations. However, often it is advantageous to exploit the following inequality.

Scheffe's identity: *For two densities f and g defined on $\mathbb{R}^m$ one has*

$$\sup_{\tilde{B}\in\mathcal{B}^m} |\int\limits_{\tilde{B}} f(z)dz - \int\limits_{\tilde{B}} g(z)dz| = \frac{1}{2}\left(\int\limits_{\mathbb{R}^m} |f(z) - g(z)|dz\right).$$

Thus we immediately obtain for a Borel set $M \subset \mathbb{R}^p$

$$\forall j \in J :\ \sup_{x\in M} |\hat{g}_0^j(x) - \hat{g}_n^j(x,\omega)| \leq \frac{1}{2}\int\limits_{\mathbb{R}^m} |f_0(z) - f_n(z)|dz.$$

Hence, under the assumptions of Theorem 3

$$P(\omega : \max_{j\in J} \sup_{x\in M} |g_0^j(x) - g_n^j(x,\omega)| \geq 2\beta_{n,\kappa}) \leq \mathcal{H}(\kappa)$$

with $\beta_{n,\kappa}$ and $\mathcal{H}$ as given in Theorem 3.

4 Proofs for Section 2

An important tool for our investigations is the bounded differences (or McDiarmid's) inequality in the multivariate form. For the readers convenience we quote this inequality. For a proof see for instance [13] or [1].

Bounded Differences Inequality. Let $Z_1, \ldots, Z_n$ be i.i.d. random vectors with values in $\mathbb{R}^m$ and $g|\mathbb{R}^m \to \mathbb{R}^1$ a measurable function.

If $\forall i = 1, \ldots, n$
$\sup\limits_{z_1,\ldots,z_n,z_i'} |g(z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_n) - g(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_n)| \le c_i$
then

$$P(g(Z_1, \ldots, Z_n) - \mathbb{E}g(Z_1, \ldots, Z_n) \ge t) \le \exp(-\frac{2t^2}{\sum\limits_{i=1}^{n} c_i^2}).$$

The difference $|\mathbb{E}f_n(z) - f_0(z)|$ plays a role in the uniform setting and in the L_1 setting as well. Hence we start with the investigation of this term. We have to distinguish whether ($Vf2$-$\mathcal{Z}$) or (VfH-$\mathcal{Z}$) is satisfied.

Lemma 1. Let (K1), (K2), (K5), (K6), and ($Vf2$-$\mathcal{Z}$) be satisfied. Then for a kernel estimator f_n with bandwidth h_n and kernel K the inequality

$$\sup_{z\in\mathcal{Z}} |\mathbb{E}f_n(z) - f_0(z)| \le \frac{C_3C_4}{2} h_n^{2/m}$$

holds.

Proof of Lemma 1. We have $\mathbb{E}f_n(z) = \frac{1}{h_n} \int\limits_{\mathbb{R}^m} K(\frac{z-y}{h_n^{1/m}}) f_0(y)dy$. Hence $\mathbb{E}f_n(z) - f_0(z) = \int_{\mathbb{R}^m} K(z-y)f_0(y)dy - f_0(z)$. Now we take into account that $K(u) = K(-u)$, change variables $z - y = u$, and exploit further properties of K:

$$\mathbb{E}f_n(z) - f_0(z) = \int_{\mathbb{R}^m} K(u)f_0(z + h_n^{1/m}u)du - f_0(z).$$

Because of $\int_{\mathbb{R}^m} K(u)du = 1$ we have $\int_{\mathbb{R}^m} K(u)f_0(z)du - f_0(z) = 0$ and hence

$$\mathbb{E}f_n(z) - f_0(z) = \int_{\mathbb{R}^m} K(u)(f_0(z + h_n^{1/m}u) - f_0(z))du.$$

The Taylor expansion of f_0 yields

$$f_0(z + h_n^{1/m}u) = f_0(z) + h_n^{1/m}u^T f_0'(z) + \frac{1}{2}h_n^{2/m}u^T H_0(\zeta_{z,u})u$$

where $f_0'(z)$ denotes the gradient, H_0 denotes the Hessian of f_0, and $\zeta_{z,u} \in [z, z + uh_n^{1/m}]$. Consequently

$$\mathbb{E}f_n(z) - f_0(z) = \int_{\mathbb{R}^m} K(u)(h_n^{1/m}u^T f_0'(z) + \frac{1}{2}h_n^{2/m}u^T H_0(\zeta_{z,u})udu.$$

Because of $\int\limits_{\mathbb{R}^m} u_iK(u)du = 0$ we obtain $|\mathbb{E}f_n(z) - f_0(z)| \le \frac{1}{2}C_3C_4h_n^{2/m}$. $\square$

Since $\frac{1}{2}C_3C_4h_n^{2/m}$ does not depend on z, we immediately obtain the following corollary.

Corollary. Under the assumptions of Lemma 1 the inequalities

$$\sup_{z\in\mathcal{Z}} |\mathbb{E}f_n(z) - f_0(z)| \leq \tfrac{1}{2}C_3C_4h_n^{2/m} \text{ and}$$
$$\int\limits_{\mathcal{T}} |\mathbb{E}f_n(z) - f_0(z)|\, dz \leq \tfrac{1}{2}C_3C_4h_n^{2/m}\lambda(\mathcal{T})$$

hold.

If we only impose the Hölder condition (VfH-$\mathcal{Z}$) we obtain Lemma 2.

Lemma 2. Let (K1), (K2), (K8), and (VfH-$\mathcal{Z}$) be satisfied. Then for a kernel estimator f_n with bandwidth h_n und kernel K the inequality

$$|\mathbb{E}f_n(z) - f_0(z)| \leq L\mu^\alpha h_n^{\alpha/m}$$

holds.

Proof of Lemma 2. We start as in the proof of Lemma 1. Instead of exploiting the Taylor expansion we proceed as follows:
$|\mathbb{E}f_n(z) - f_0(z)| \leq \int_{\mathbb{R}^m} |K(u)|L||h_n^{1/m}u||^\alpha du = L\mu^\alpha h_n^{\alpha/m}$. □

Lemma 2 yields the following corollary.

Corollary. Under the assumptions of Lemma 2 the inequalities

$$\sup_{z\in\mathcal{Z}} |\mathbb{E}f_n(z) - f_0(z)| \leq L\mu^\alpha h_n^{\alpha/m} \text{ and}$$
$$\int\limits_{\mathcal{T}} |\mathbb{E}f_n(z) - f_0(z)|\, dz \leq L\mu^\alpha h_n^{\alpha/m}\lambda(\mathcal{T})$$

hold.

Proof of Theorem 1.
We proceed as follows. Obviously, $\sup\limits_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| \leq T1_n + T2_n + T3_n$ where

$$\begin{aligned}
T1_n &:= \left|\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| - \mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)|\right]\right|, \\
T2_n &:= \mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|\right], \\
T3_n &:= \sup_{z\in\mathcal{Z}} |\mathbb{E}f_n(z) - f_0(z)|\,.
\end{aligned}$$

$T3_n$ was considered in the foregoing corollaries. In the next step we investigate the random part $T1_n$.

Lemma 3. Let the condition (K3) be satisfied.Then for a kernel estimator f_n with bandwidth h_n and kernel K the following inequality holds:

$$\forall \kappa > 0: \; P\left(\left| \sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| - \mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)|\right] \right| \geq \frac{\kappa}{\sqrt{n}h_n} \right) \leq 2e^{-\frac{2\kappa^2}{C_1{}^2}}.$$

Proof of Lemma 3. Let $g(z_1, ..., z_n) := \sup_{z\in\mathcal{Z}} \left| \tilde{f}_n(z, z_1, ..., z_n) - f_0(z) \right|$ where $\tilde{f}_n(z, z_1, ..., z_n) := \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{z-z_i}{h_n^{1/m}}\right)$.
Then we have

$$\begin{aligned}
&\sup_{z_1,...,z_n \wedge z_i' \in A} |g(z_1, ..., z_n) - g(z_1, ..., z_i', ..., z_n)| \\
&= \sup_{z_1,...,z_n \wedge z_i' \in \mathbb{R}^m} \left| \sup_{z\in\mathcal{Z}} \left| \tilde{f}_n(z, z_1, ..., z_n) - f_0(z) \right| - \sup_{z\in\mathcal{Z}} \left| \tilde{f}_n(z, z_1, ..., z_i', ..., z_n) - f_0(z) \right| \right| \\
&\leq \sup_{z_1,...,z_n \wedge z_i' \in \mathbb{R}^m} \sup_{z\in\mathcal{Z}} \left| \tilde{f}_n(z, z_1, ..., z_i, ..., z_n) - \tilde{f}_n(z, z_1, ..., z_i', ..., z_n) \right| \\
&= \sup_{z_1,...,z_n \wedge z_i' \in \mathbb{R}^m} \sup_{z\in\mathcal{Z}} \Big| \tfrac{1}{nh_n} \left(K(\tfrac{z-z_1}{h_n^{1/m}}) + ... + K(\tfrac{z-z_i}{h_n^{1/m}}) + ... + K(\tfrac{z-z_n}{h_n^{1/m}}) \right) \\
&\qquad - \tfrac{1}{nh_n} \left(K(\tfrac{z-z_1}{h_n^{1/m}}) + ... + K(\tfrac{z-z_i'}{h_n^{1/m}}) + ... + K(\tfrac{z-z_n}{h_n^{1/m}}) \right) \Big| \\
&= \sup_{z_1,...,z_n \wedge z_i' \in \mathbb{R}^m} \sup_{z\in\mathcal{Z}} \left| \tfrac{1}{nh_n} (K(\tfrac{z-z_i}{h_n^{1/m}}) - K(\tfrac{z-z_i'}{h_n^{1/m}})) \right|.
\end{aligned}$$

According to (K3) we obtain

$$\left| K(\frac{z-z_i}{h_n^{1/m}}) - K(\frac{z-z_i'}{h_n^{1/m}}) \right| \leq C_1.$$

Consequently the assumption of the bounded differences inequality is satisfied with $c_i := \frac{1}{nh_n} C_1$, $i = 1,...n$, and we have

$\forall\, t > 0: \; P(|g(Z_1, ..., Z_n) - \mathbb{E}g(Z_1, ..., Z_n)| \geq t) \leq 2e^{-\frac{2t^2 nh_n^2}{C_1{}^2}}$ and
$\forall\, \kappa > 0: \; P(|g(Z_1, ..., Z_n) - \mathbb{E}g(Z_1, ..., Z_n)| \geq \frac{\kappa}{\sqrt{n}h_n}) \leq 2e^{-2\frac{\kappa^2}{C_1{}^2}}$. □

For the investigation of the term $T2_n$ we use the Fourier transform k of the kernel K, see Section 2. Furthermore, we need the Fourier transform of f_0 :

$$\phi(u) := \int_{\mathbb{R}^m} e^{iu^T y} f_0(y) dy \quad \forall\, u \in \mathbb{R}^m.$$

Lemma 4. Let (K2), (K4), and (K7) be satisfied. Then for a kernel estimator f_n with bandwidth h_n and kernel K the following inequality holds

$$\mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|\right] \leq \frac{C_2}{(2\pi)^m \sqrt{n} h_n}.$$

Proof of Lemma 4. Because of (K7) we have $\int_{\mathbb{R}^m} |k(u)|du < \infty$. Hence we can employ the inversion formula and obtain $K(u) := (\frac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T y} k(y)dy \quad \forall\, u \in \mathbb{R}^m$. Consequently f_n can be rewritten in the following form:

$$
\begin{aligned}
f_n(z) &= \frac{1}{nh_n} \sum_{l=1}^{n} \left((\frac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-i\left(\frac{z-Z_l}{h_n^{1/m}}\right)^T y} k(y)dy \right) \\
&= \frac{1}{nh_n} \sum_{l=1}^{n} \left((\frac{1}{2\pi})^m h_n \int_{\mathbb{R}^m} e^{-i(z-Z_l)^T u} k(h_n^{1/m}u)du \right) \\
&= (\frac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iz^T u} k(h_n^{1/m}u) \left(\frac{1}{n} \sum_{l=1}^{n} e^{iZ_l^T u} \right) du
\end{aligned}
$$

With $\phi_n(u) := \frac{1}{n}\sum_{l=1}^n e^{iZ_l^T u} \quad \forall\, u \in \mathbb{R}^m$ we have

$$
f_n(z) = (\frac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iz^T u} k(h_n^{1/m}u)\phi_n(u)du \quad \forall\, z \in \mathcal{Z}.
$$

Because of Jensen's inequality we obtain

$$
\mathbb{E}^2 \left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)| \right] \le \mathbb{E} \left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|^2 \right].
$$

Now we use the Fourier transform and employ Fubini's theorem:

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|^2\right] \\
&= \mathbb{E}\left[\sup_{z\in\mathcal{Z}} \left|(\tfrac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T z} k(h_n^{1/m}u)\phi_n(u)du - \mathbb{E}\left[(\tfrac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T z} k(h_n^{1/m}u)\phi_n(u)du\right]\right|^2\right] \\
&= \mathbb{E}\left[\sup_{z\in\mathcal{Z}} \left|(\tfrac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T z} k(h_n^{1/m}u)\phi_n(u)du - (\tfrac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T z} k(h_n^{1/m}u)\mathbb{E}\phi_n(u)du\right|^2\right] \\
&= \mathbb{E}\left[\sup_{z\in\mathcal{Z}} \left|(\tfrac{1}{2\pi})^m \int_{\mathbb{R}^m} e^{-iu^T z} k(h_n^{1/m}u)(\phi_n(u) - \mathbb{E}\phi_n(u))du\right|^2\right].
\end{aligned}
$$

With $\left|e^{-iu^T z}\right| = 1$ we can conclude that

$$
\mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|^2\right] \le \mathbb{E}\left[\left|(\frac{1}{2\pi})^m \int_{\mathbb{R}^m} \left|k(h_n^{1/m}u)\right| |\phi_n(u) - \mathbb{E}\phi_n(u)|\, du\right|^2\right].
$$

For an integrable real-valued function $(u,\omega) \to \tilde{z}(u,\omega) =: z(u),\ u \in \mathbb{R}^m$, we have because of Minkowski's integral inequality

$$
\begin{aligned}
\mathbb{E}\left[(\int_{\mathbb{R}^m} z(u)du)^2\right] &= \mathbb{E}\left[(\int_{\mathbb{R}^m} z(u)du)(\int_{\mathbb{R}^m} z(v)dv)\right] \\
&\le \left(\int_{\mathbb{R}^m} \mathbb{E}^{\frac{1}{2}}[z^2(u)]du\right)^2.
\end{aligned}
$$

Hence with $z(u) = \left|k(h_n^{1/m}u)\right| |\phi_n(u) - \mathbb{E}\phi_n(u)|$ we obtain

$$\mathbb{E}^{\frac{1}{2}}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|^2\right] \le (\frac{1}{2\pi})^m \int_{\mathbb{R}^m} \mathbb{E}^{\frac{1}{2}}\left[\left|k(h_n^{1/m}u)\right|^2 |\phi_n(u) - \mathbb{E}\phi_n(u)|^2\right] du.$$

Furthermore

$$\mathbb{E}^{\frac{1}{2}}\left[\left|k(h_n^{1/m}u)\right|^2 |\phi_n(u) - \mathbb{E}\phi_n(u)|^2\right]$$
$$= \left|k(h_n^{1/m}u)\right| \mathbb{E}^{\frac{1}{2}}\left[|\phi_n(u) - \mathbb{E}\phi_n(u)|^2\right] = \left|k(h_n^{1/m}u)\right| \sqrt{\mathrm{var}(\phi_n(u))},$$

where $\mathrm{var}\,(\phi_n(u)) = \frac{1}{n^2}\sum_{k=1}^n \mathrm{var}\left(e^{iu^T Z_k}\right) = \frac{1}{n}\mathrm{var}\left(e^{iu^T Z_1}\right) \le \frac{1}{n}$.

Summarizing,

$$\begin{aligned}
\mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|\right] &\le \mathbb{E}^{\frac{1}{2}}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|^2\right] \\
&\le (\frac{1}{2\pi})^m \int_{\mathbb{R}^m} \left|k(h_n^{1/m}u)\right| \sqrt{\mathrm{var}(\phi_n(u))} du \\
&\le (\frac{1}{2\pi})^m \frac{1}{\sqrt{n}} \int_{\mathbb{R}^m} \left|k(h_n^{1/m}u)\right| du \\
&= (\frac{1}{2\pi})^m \frac{1}{\sqrt{n}} \frac{1}{h_n} \int_{\mathbb{R}^m} |k(y)|\, dy.
\end{aligned}$$

With the assumption (K4) the inequality

$$\mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - \mathbb{E}f_n(z)|\right] \le \frac{C_2}{(2\pi)^m \sqrt{n} h_n}.$$

follows. □

From the triangle inequality, Lemma 3, and Lemma 4 we have

$\sup_{n\in\mathbb{N}} P(\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| \ge \beta_{n,\kappa}) \le \sup_{n\in\mathbb{N}} P(T1_n + T2_n + T3_n \ge \beta_{n,\kappa}).$
$\le \sup_{n\in\mathbb{N}} P\left(T1_n + \frac{C_2}{(2\pi)^m\sqrt{n}h_n} + T3_n \ge \beta_{n,\kappa}\right)$
$= \sup_{n\in N} P\left(|\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)| - \mathbb{E}\left[\sup_{z\in\mathcal{Z}} |f_n(z) - f_0(z)|\right]| \ge \frac{\kappa}{\sqrt{n}h_n}\right)$
$\le 2e^{\frac{-2\kappa^2}{{C_1}^2}}.$

Inserting the special forms of $T3_n$ from the corollaries we obtain the assertion of Theorem 1. □

Proof of Theorem 2.
We start with the following inequality for a fixed $z \in \mathbb{R}^m$.

$$|f_n(z) - f_0(z)| \le |f_n(z) - \mathbb{E}f_n(z)| + |\mathbb{E}f_n(z) - f_0(z)|\,.$$

Now the bounded differences inequality is applied to $g(z_1, \ldots, z_n) = \tilde{f}_n(z, z_1, \ldots, z_n)$ where $\tilde{f}_n$ is defined as in the proof of Lemma 3. We obtain similarly as in the first part of the proof to Theorem 1

$$P\left(|f_n(z) - \mathbb{E}f_n(z)| \ge \frac{\kappa}{\sqrt{n}h_n}\right) \le 2e^{-2\frac{\kappa^2}{C_1^2}}.$$

For the second term in the above inequality we use Lemma 1 or Lemma 2. □

Proof of Theorem 3.
We start with the triangle inequality in the following form:

$\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz \le U1_n + U2_n + U3_n$ where

$$\begin{aligned} U1_n &:= \left| \int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz - \mathbb{E}\left[\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz\right]\right|, \\ U2_n &:= \mathbb{E}\left[\int\limits_{\mathcal{T}} |f_n(z) - \mathbb{E}f_n(z)|\, dz\right], \\ U3_n &:= \int\limits_{\mathcal{T}} |\mathbb{E}f_n(z) - f_0(z)|\, dz. \end{aligned}$$

Firstly we investigate the random part $U1_n$. In the bounded differences inequality we use $g(z_1, \ldots, z_n) = \int\limits_{\mathcal{T}} |\tilde{f}_n(z, z_1, \ldots, z_n) - f_0(z)| dz$ with $\tilde{f}_n$ as in the proof of Theorem 1 and obtain:
$|g(z_1, \ldots, z_i, \ldots, z_n) - g(z_1, \ldots, z_i', \ldots, z_n)| \le \frac{1}{n} \int\limits_{\mathcal{T}} |K(z - z_i) - K(z - z_i')| dz.$

Hence, with $c_i = \frac{2}{n}C_5$, the bounded differences inequality yields

$$P\left(\left|\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz - \mathbb{E}\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz\right| \ge t\right) \le e^{\frac{-nt^2}{2C_5^2}} \quad \text{and}$$
$$P\left(\left|\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz - \mathbb{E}\int\limits_{\mathcal{T}} |f_n(z) - f_0(z)|\, dz\right| \ge \tfrac{\kappa}{\sqrt{n}}\right) \le e^{\frac{-\kappa^2}{2C_5^2}}.$$

Now we can proceed as in the second part of the proof to Theorem 9.5 in [3], i.e., we take into account that

$$\mathbb{E}\int\limits_{\mathcal{T}} |f_n(z) - \mathbb{E}f_n(z)| dz = \mathbb{E}\int\limits_{\mathcal{T}} |f_n(z) - \int\limits_{\mathbb{R}^m} K(z - y) f_0(y) dy| dz,$$

use Fubini's theorem, employ $\mathbb{E}|X| \leq \sqrt{\mathbb{E}(X^2)}$ for a random variable X, and utilize finally the Cauchy Schwarz inequality and Young's inequality. We obtain

$$\mathbb{E} \int_{\mathcal{T}} |f_n(z) - \mathbb{E} f_n(z)| dz \leq \frac{1}{\sqrt{n h_n}} \sqrt{\lambda(U_{h_n} \mathcal{T}) C_6}.$$

It remains to apply the corollaries for $U3_n$. □

Acknowledgement. The authors would like to thank Klaus Ziegler for valuable discussions about the topic.

References

[1] Bartlett, Peter (2006). " Proof of the bounded differences inequality." www.cs.berkeley.edu/ bartlett/.../bdddiff.pdf

[2] Bayraksan, Guzin, and David P. Morton, 2011. "A sequential sampling procedure for stochastic programming." *Operations Research* 59:898–913.

[3] Devroye, Luc, and Lugosi, Gábor (2001). *Combinatorial Methods in Density Estimation.* Springer.

[4] Dupačová, Jitka (1986). " Stability in stochastic programming with recourse-contaminated distributions." *Math. Programming Study* 27:133–144.

[5] Dupačová, Jitka, and Milos Kopa (2012). "Robustness in stochastic programs with risk constraints." *Annals of Oper. Research* 200:55–74.

[6] Dupačová, Jitka, and Roger J-B Wets (1988). "Asymptotic behavior of statistical estimators and of optimal solutions of stochastic problems." *Ann. Statist.* 16:1517–1549.

[7] Dony, Julia, and Uwe Einmahl (2009). " Uniform in bandwidth consistency of kernel estimators at a fixed point." In: *High Dimensional Probability V*, edited by Christian Houdré et al., 308–325.

[8] Dünnbier, A. (2008). "Konfidenzintervalle für Modalwerte." Diploma thesis, TU Ilmenau.

[9] Gersch, Oliver. 2006. "Convergence in Distribution of Random Closed Sets and Applications in Stability Theory of Stochastic Optimization." PhD Thesis, TU Ilmenau.

[10] Hanasusanto, Grani Adiwena, Vladimir Roitch, Daniel Kuhn, and Wolfram Wiesemann (2016). " Ambiguous joint chance constraints under mean and dispersion information." *Operations Research.*

[11] Mason, David, M., and Wolfgang Polonik (2009). “Asymptotic normality of plug-in level set estimates.” *Ann. Appl. Probab.* 19:1108–1142.

[12] Müller, Hans-Georg (1984). “Smooth optimum kernel estimators of densities, regression curves and modes.” *Ann. Math. Statist.* 12:766 –774.

[13] McDiarmid, Colin (1989). “On the method of bounded differences.” In: *Surveys in Combinatorics*, edited by Johannes Siemons, 148 –188, Cambridge Univ. Press.

[14] Pagnocelli, Bernardo K., Shabbir Ahmed, and Alexander Shapiro (2009). “Sample average approximation method for chance constrained programming: Theory and applications.” *J. Optim. Theory Appl.* 142:399–416.

[15] Parzen, Emanuel (1962). “On estimation of a probability density function and mode.” *Ann. Math. Statist.* 33:1065 –1076.

[16] Pflug, Georg Ch. (1992). “Asymptotic dominance and confidence for solutions of stochastic programs.” *Czechoslovak J. for Operations Research* 1:21–30.

[17] Pflug, Georg Ch. (2003). “Stochastic Optimization and Statistical Inference.” In: *Stochastic Programming*, edited by Andrzej Ruszczyński and Alexander Shapiro, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 427–482.

[18] Pflug, Georg Ch., and Alois Pichler (2014). *Multistage Stochastic Optimization.* Springer Series in Operations Research and Financial Engineering

[19] Römisch, Werner (2003). “Stability of stochastic programming.” In *Stochastic Programming*, edited by Andrzej Ruszczyński and Alexander Shapiro, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 483–554.

[20] Römisch, Werner, and Rüdiger Schultz (1991). “Distribution sensitivity in stochastic programming.” *Math. Programming* 50:197–226.

[21] Rosenblatt, Murray (1956). “Remarks on some non-parametric estimates of a density function.” *Ann.Math. Statist.* 27:832–837.

[22] Royset, Johannes O., Roger J-B Wets (2016). “Multivariate Epi-splines and evolving function identification problems.” *Set-valued Var. Anal.* 24:517–545.

[23] Shapiro, Alexander (2003). “Monte Carlo sampling methods.” In *Stochastic Programming*, edited by Andrzej Ruszczyński and Alexander Shapiro, Handbooks in Operations Research and Management Science, Vol. 10, Elsevier, 353—425.

[24] Shapiro, Alexander, and Shabbir Ahmed (2004). "On a class of minimax stochastic programs." *SIAM J. Optim.* 14:1237–1249.

[25] Seeger, Stefan (2012). "Kerndichteschätzer im Kontext universaler Konfidenz." Master-Thesis, TU Ilmenau.

[26] Silverman, Bernard W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.

[27] Sinotina, Tatiana, and Silvia Vogel. (2012). "Universal confidence sets for the mode of a regression function." *IMA Journal of Management Mathematics* 23:309–323.

[28] Vogel, Silvia (2005). "Qualitative stability of stochastic programs with applications in asymptotic statistics." *Statistics and Decisions* 23:219–248.

[29] Vogel, Silvia (2008). "Universal confidence sets for solutions of optimization problems." *SIAM J. on Optimization* 19:1467–1488.

[30] Vogel, Silvia (2008). "Confidence sets and convergence of random functions." In: *Festschrift in Celebration of Prof. Dr. Wilfried Grecksch's 60th Birthday*, edited by Christiane Tammer and Frank Heyde, Shaker-Verlag.

[31] Wieczorek, Barbara, and Klaus Ziegler (2010). "On optimal estimation of a non-smooth mode in a nonparametric regression model with [alpha]-mixing errors." *Journal of Statistical Planning and Inference* 140:406–418.

[32] Xu, Chuan, Cheng, Jianqing Cheng, and Abdel Lisser (2015). "A sampling method to chance-constrained semidefinite optimization." *ICORES 2015*, 75–81.

[33] Žačková, Jitka (1966). " On minimax solutions of stochastic linear programming problems." *Časopis pro Pěstování Matematiky* 91:423-430.