



HAL
open science

Nearest neighbor based conformal prediction

Laszlo Gyôrfi, Harro Walk

► **To cite this version:**

Laszlo Gyôrfi, Harro Walk. Nearest neighbor based conformal prediction. Annales de l'ISUP, 2019, 63 (2-3), pp.173-190. hal-03603867

HAL Id: hal-03603867

<https://hal.science/hal-03603867v1>

Submitted on 10 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris

63, fasc. 2-3, 2019, 173-190

Numéro spécial en l'honneur des 80 ans de Denis Bosq /

Special issue in honour of Denis Bosq's 80th birthday

Nearest neighbor based conformal prediction

László Györfi^{*,†}, and Harro Walk[‡]

Budapest University of Technology and Economics[†] and University of Stuttgart[‡]

Abstract: In this paper we introduce a nearest neighbor based estimate of the prediction interval with prescribed conditional coverage probability and with small length. In the special case, when there is no feature vector, the problem is the estimate of a confidence interval. For confidence interval estimate, we show the distribution-free strong consistency of the conditional coverage probability and of excess length of the interval, while the conditional coverage probability of prediction interval has the distribution-free strong consistency property and under weak conditions on the underlying distributions strong consistency and the fast rate of convergence of the excess length are shown. As a consequence, we construct a confidence set estimate for classification.

1. Introduction

Consider a training data set

$$(1.1) \quad \mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\},$$

and a test point (X_{n+1}, Y_{n+1}) , with the training and test data all drawn i.i.d. from the same distribution. Here each $X_i \in \mathbb{R}^d$ is a feature vector, while $Y_i \in \mathbb{R}$ is a response variable. The problem of predictive inference is the following: if we observe the n training data points \mathcal{D}_n , and are given the feature vector X_{n+1} for a new test data point, we would like construct a prediction interval $C_n(x) = C_n(\mathcal{D}_n, x)$ for Y_{n+1} such that we believe $C_n(X_{n+1})$ is likely to contain the test point's true response value Y_{n+1} .

We are interested in the conditional coverage probability

$$(1.2) \quad \mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\}$$

and in the length of the interval $C_n(x)$. Most of the related literature study the (unconditional) coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1})\},$$

while Vovk [16] considers the conditional coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n\}$$

^{*}Corresponding author

AMS 2000 subject classifications: Primary 62G08, 62G20

Keywords and phrases: confidence interval, prediction interval, conditional coverage probability, excess length, classification, nearest neighbor estimate

and Barber et al. [1] investigate the conditional coverage probability

$$\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid X_{n+1} = x\}.$$

If there is no feature vector X_i , then the problem is reduced to the estimation of confidence interval.

In this paper we consider three set estimation problems:

- confidence interval of given conditional coverage probability and of small length,
- prediction interval of given conditional coverage probability (1.2) and of small length,
- confidence set for classification with given conditional coverage probability (1.2) and with small size.

Such estimates can be derived from the estimates of quantiles or conditional quantiles. For example, concerning the confidence interval one may search for the shortest interval with prescribed empirical distribution. In the sequel we modify this principle such that choose the shortest interval with a bit smaller empirical distribution. It turns out that these interval estimation problems are easier than the corresponding point estimation, which means that the rate of convergence of the length of the interval is super-polynomial.

Lei et al. [12], Papadopoulos et al. [14], Romano, Patterson and Candés [15], Vovk [16], studied estimates of prediction intervals. These authors discuss and investigate various concepts on the basis of independent and identically distributed or exchangeable samples. Concerning an overview on the validity and on efficiency, especially a negative result for the non-asymptotic situation see Vovk [16] and Barber et al. [1]. Lei and Wasserman [11] and Lei, Robins and Wasserman [13] contain results on batch mode prediction, while Vovk [16] and Lei et al. [12] are on online (or sequential) prediction.

Usually, the authors use a prediction set around a point prediction of Y , especially intervals are considered, which are symmetric around a regression function estimate, showing the consistency of the coverage probability, also refining the procedures under computational aspects (see Papadopoulos et al. [14]). Here we show the consistency of the conditional coverage probability and investigate the rate of convergence for the excess of the estimated prediction (confidence) interval length with respect to the minimal prediction (confidence) interval length.

2. Estimate the confidence interval

In this setup there is no feature vector. For the random variable Y , introduce the distribution function

$$F(y) := \mathbb{P}\{Y \leq y\}.$$

If $0 < p < 1$, then the set of possible quantiles is an interval Q such that

$$[q_{p,low}, q_{p,up}) \subset Q \subset [q_{p,low}, q_{p,up}],$$

where

$$q_{p,up} := F_{up}^{-1}(p) := \sup_{y \in \mathbb{R}: F(y) \leq p} y$$

is the upper quantile, and

$$q_{p,low} := F_{low}^{-1}(p) := \min_{y \in \mathbb{R}: F(y) \geq p} y$$

denotes the lower quantile. In general, the quantile is not unique. The quantile is unique, if

$$q_{p,low} = q_{p,up}.$$

Concerning a confidence interval

$$C = [f_{lo}, f_{hi}],$$

we need that for the *coverage probability*

$$\mathbb{P}\{Y \in C\} \geq 1 - \alpha,$$

where $0 < \alpha < 1$ is fixed, and the *size of the interval*

$$\Delta := f_{hi} - f_{lo}$$

is as small as possible.

If the lower and upper quantiles are known, then the optimal confidence interval can be constructed as follows: Put

$$D(p) := F_{low}^{-1}(p + 1 - \alpha) - F_{up}^{-1}(p),$$

($0 < p < \alpha$) and

$$p^* := \arg \min_{0 < p < \alpha} D(p).$$

Then

$$\Delta^* = \min_{0 < p < \alpha} D(p) = D(p^*),$$

$$f_{lo}^* = F_{up}^{-1}(p^*)$$

and

$$f_{hi}^* = F_{low}^{-1}(p^* + 1 - \alpha).$$

For the sake of simplicity, throughout the paper we assume that the $\arg \min$ exists. In the general case, p^* can be defined such that $D(p^*) \leq \inf_{0 < p < \alpha} D(p) + 1/n^2$, where n is the sample size.

If the lower and upper quantiles are unknown, then assume, that we observed data $\{Y_1, \dots, Y_n\}$ consisting of independent and identically distributed copies of Y .

Let

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{Y_i \leq y},$$

be the empirical distribution function, where \mathbb{I} denotes the indicator function, while

$$F_{n,low}^{-1}(p) := \min_{y \in \mathbb{R}: F_n(y) \geq p} y, \quad 0 < p < 1$$

and

$$F_{n,up}^{-1}(p) := \sup_{y \in \mathbb{R}: F_n(y) \leq p} y, \quad 0 < p < 1$$

are the corresponding quantile function estimates. From these estimates one may derive the plug-in estimate of the optimal confidence interval. Put

$$D_n(p) := F_{n,low}^{-1}(p + 1 - \alpha - \ln n / \sqrt{n}) - F_{n,up}^{-1}(p + \ln n / \sqrt{n})$$

($0 < p < \alpha$) and

$$p_n := \arg \min_{0 < p < \alpha} D_n(p).$$

Then

$$\Delta_n = D_n(p_n),$$

$$f_{n,lo} = F_{n,up}^{-1}(p_n + \ln n / \sqrt{n})$$

and

$$f_{n,hi} = F_{n,low}^{-1}(p_n + 1 - \alpha - \ln n / \sqrt{n}).$$

The main point here is that we underestimate the optimal confidence interval resulting in distribution-free consistency.

Lei, Robins and Wasserman [13] studied the extension of our problem, when Y is d' dimensional and the aim was to approximate the confidence set of minimal Lebesgue measure. They assumed a smooth density of Y , and from the kernel density estimate they derived a confidence set estimate. For Lipschitz continuous density and for the expected Lebesgue measure of the symmetric difference of the optimal and estimated confidence set they got a rate of convergence $O(\ln n / n^{1/(d'+2)})$, which is $O(\ln n / n^{1/3})$, for $d' = 1$. Here we do not assume a density and construct a confidence interval instead of a confidence set such that the tail of the excess of the estimated confidence interval length with respect to the minimal confidence interval has the rate of convergence $O(1/n^{2 \ln n})$.

Put

$$C_n = [f_{n,lo}, f_{n,hi}].$$

The following theorem is on the tail distribution of the conditional coverage probability $\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\}$ and of the excess length $(\Delta_n - \Delta^*)^+$:

Theorem 2.1. For any distribution of Y and $\varepsilon > 0$,

$$(2.1) \quad \begin{aligned} & \mathbb{P}\{|\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} - (1 - \alpha)| \geq \varepsilon\} \\ & \leq 8(n+1)e^{-n\varepsilon^2/128} + \mathbb{I}_{\ln n/\sqrt{n} \geq \varepsilon/4}, \end{aligned}$$

and

$$(2.2) \quad \mathbb{P}\{\Delta_n - \Delta^* \geq 0\} \leq 2/n^{2 \ln n}.$$

The proof of this theorem is in the last section. (2.1) implies that

$$(2.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} = 1 - \alpha$$

a.s. and

$$(2.4) \quad \mathbb{E}\{|\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} - (1 - \alpha)|\} = O(\ln n/\sqrt{n}).$$

Furthermore, because of

$$\sum_{n=1}^{\infty} \mathbb{P}\{\Delta_n - \Delta^* \geq 0\} \leq \sum_{n=1}^{\infty} 2/n^{2 \ln n} < \infty,$$

the Borel-Cantelli lemma implies the strong distribution-free consistency of the excess length:

$$\lim_{n \rightarrow \infty} [\Delta_n - \Delta^*]^+ = 0$$

a.s.

3. Conformal prediction

Introduce the conditional distribution function

$$F(y \mid x) := \mathbb{P}\{Y \leq y \mid X = x\}$$

and the corresponding lower and upper conditional quantile functions are defined by

$$q_{p,low}(x) := F_{x,low}^{-1}(p) := \min_{y \in \mathbb{R}: F(y|x) \geq p} y,$$

and

$$q_{p,up}(x) := F_{x,up}^{-1}(p) := \sup_{y \in \mathbb{R}: F(y|x) \leq p} y,$$

respectively.

The conformal prediction is an interval

$$C(x) = [f_{lo}(x), f_{hi}(x)].$$

Concerning a prediction interval $C(x)$, we need that for the *conditional coverage probability* (called also conditional validity in [11] and in [16])

$$(3.1) \quad \mathbb{P}\{Y \in C(X) \mid X = x\} \geq 1 - \alpha,$$

where $0 < \alpha < 1$ is fixed, and the *length of the interval*

$$\Delta(x) := f_{hi}(x) - f_{lo}(x)$$

is as small as possible (efficiency, compare [11] and [16]).

An extension of our problem can be found in Lei et al. [12] (with a generalization to multi-dimensional Y in Lei and Wasserman [11]) such that $C(x)$ is an arbitrary (measurable) set, and under (3.1) the aim is to minimize $\lambda(C(x))$, where λ stands for the Lebesgue measure. From a kernel density estimate they derived a consistent estimate of the optimal prediction set, and for Lipschitz continuous conditional density showed the rate of convergence of order $O(\ln n/n^{1/(d+3)})$. Our setup is much simpler such that there is no need to assume the existence of the density of (X, Y) . Here, we do not assume a density of (X, Y) . Interestingly, under mild conditions, especially Lipschitz continuity of the function $F(y \mid \cdot)$, we get a super-polynomial rate $O(1/n^{2 \ln n})$ in the case of prediction interval, instead of a prediction set.

If the conditional distribution function and the conditional quantile functions are known, then the optimal prediction interval can be constructed. Put

$$D(p) := F_{x,low}^{-1}(p + 1 - \alpha) - F_{x,up}^{-1}(p),$$

($0 < p < \alpha$) and

$$p^* := p^*(x) := \arg \min_{0 < p < \alpha} D(p).$$

Then

$$\Delta^*(x) = \min_{p < \alpha} D(p) = D(p^*),$$

$$f_{lo}^*(x) = F_{x,up}^{-1}(p^*)$$

and

$$f_{hi}^*(x) = F_{x,low}^{-1}(p^* + 1 - \alpha).$$

If the conditional distribution function and the conditional quantile function are unknown, then assume, that we observed data (1.1). The obvious way is to estimate the conditional distribution function and the conditional quantile function. Allowing additional measurement errors, Hansmann and Kohler [10] studied local averaging regression based, in detail kernel regression based, conditional quantile function estimates. Here we introduce a k-nearest-neighbor (k-NN) estimate, which is easier to analyze. For a fixed $x \in \mathbb{R}^d$, reorder the data $(X_1, Y_1), \dots, (X_n, Y_n)$ according to increasing values of $\|X_i - x\|$. The reordered data sequence is denoted by

$$(X_{(n,1)}(x), Y_{(n,1)}(x)), \dots, (X_{(n,n)}(x), Y_{(n,n)}(x)).$$

$X_{(n,k)}(x)$ is the k -th nearest neighbor of x . The tie breaking is done by randomization. If the distribution of feature vector X is denoted by μ , then for the sake of simplicity, we assume that μ has a density, therefore tie happens with probability 0.

The k -NN estimate of the conditional distribution function is defined by

$$F_{k,n}(y | x) = \frac{1}{k} \sum_{i=1}^k \mathbb{I}_{Y_{(n,i)}(x) \leq y},$$

while

$$F_{x,k,n,low}^{-1}(p) := \min_{y \in \mathbb{R}: F_{k,n}(y|x) \geq p} y, \quad 0 < p < 1$$

and

$$F_{x,k,n,up}^{-1}(p) := \sup_{y \in \mathbb{R}: F_{k,n}(y|x) \leq p} y, \quad 0 < p < 1$$

are the corresponding conditional quantile function estimates. From these estimates one may derive the plug-in estimate of the optimal prediction interval. Put

$$D_n(p) := F_{x,k,n,low}^{-1}(p + 1 - \alpha - t_{k,n}) - F_{x,k,n,up}^{-1}(p + t_{k,n})$$

($0 < p < \alpha$) with

$$t_{k,n} := \ln n(1/\sqrt{k} + (k/n)^{1/d}),$$

and

$$p_n := p_n(x) := \arg \min_{0 < p < \alpha} D_n(p).$$

Then

$$\Delta_n(x) = D_n(p_n),$$

$$f_{k,n,lo}(x) = F_{x,k,n,up}^{-1}(p_n + t_{k,n})$$

and

$$f_{k,n,hi}(x) = F_{x,k,n,low}^{-1}(p_n + 1 - \alpha - t_{k,n}).$$

This estimate has a simple interpretation. If $y_1 < \dots < y_k$ denotes the ordered samples of $Y_{(n,1)}(x), \dots, Y_{(n,k)}(x)$, then

$$F_{k,n}(y | x) = \frac{1}{k} \sum_{i=1}^k \mathbb{I}_{y_i \leq y}.$$

For

$$\begin{aligned} y_{g(n)} &= \arg \min_{y_i: i/k \leq \alpha} [F_{x,k,n,low}^{-1}(i/k + 1 - \alpha - 2t_{k,n}) - y_i] \\ &= \arg \min_{y_i: i \leq k\alpha} [y_{\lfloor k(i/k + 1 - \alpha - 2t_{k,n}) \rfloor} - y_i] \\ &= \arg \min_{y_i: i \leq k\alpha} [y_{\lfloor i + k(1 - \alpha - 2t_{k,n}) \rfloor} - y_i], \end{aligned}$$

we have that

$$f_{k,n,lo}(x) = y_{g(n)+\lfloor kt_{k,n} \rfloor}$$

and

$$f_{k,n,hi}(x) = y_{g(n)+\lfloor k(1-\alpha)-kt_{k,n} \rfloor}.$$

Put

$$C_{k,n}(x) = [f_{k,n,lo}(x), f_{k,n,hi}(x)].$$

The next theorem is on the strong universal consistency of the conditional coverage probability and on the rate of convergence under some smoothness condition. In order to have non-trivial rate of convergence of the conditional coverage probability, one has to assume tail and smoothness conditions, otherwise the rate of convergence can be arbitrarily slow, see Chapter 3 in [9]. For most of the related results, the feature vector X is assumed to be bounded, which excludes the classical parametric problem, where the conditional distributions of X given Y are multidimensional Gaussian distributions. Next, we introduce a mild combined tail and smoothness condition, under which we get fast rate of convergence.

Definition 3.1. For each y , the function $F(y | \cdot)$ satisfies the modified Lipschitz condition, if there is a constant C^* such that for any $y \in \mathbb{R}$ and $x, z \in \mathbb{R}^d$

$$(3.2) \quad |F(y | x) - F(y | z)| \leq C^* \mu(S_{x, \|x-z\|})^{1/d},$$

where $S_{x,r}$ denotes the sphere centered at x and having the radius r (cf. [3], [8]).

If the density f of μ is continuous, then the right hand side of (3.2) is approximately equal to $C^* f(x)^{1/d} \|x - z\|$. For $d = 1$, the standard exponential distribution is an example, where the right hand side of (3.2) is approximately equal to $C^* e^{-x} \max\{x - z, z - x\}$. Interestingly, one can show densities for modified Lipschitz condition, where $\mathbb{E}\{\|X\|\} = \infty$, for example, f is the Cauchy density.

Theorem 3.1. Assume that μ has a density. If $k = k_n$ such that $k_n/(\ln n)^2 \rightarrow \infty$ and $(\ln n)^d k_n/n \rightarrow 0$, then for any distribution of (X, Y) ,

$$(3.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{Y_{n+1} \in C_{k_n, n}(X_{n+1}) | \mathcal{D}_n, X_{n+1} = x\} = 1 - \alpha$$

a.s. for μ -almost-all x . If, in addition, for each y , the function $F(y | \cdot)$ satisfies the modified Lipschitz condition, then for any fixed x ,

$$(3.4) \quad \begin{aligned} & \mathbb{E}\{|\mathbb{P}\{Y_{n+1} \in C_{k_n, n}(X_{n+1}) | \mathcal{D}_n, X_{n+1} = x\} - (1 - \alpha)|\} \\ & = O\left(\ln n / \sqrt{k_n}\right) + O\left(\ln n (k_n/n)^{1/d}\right). \end{aligned}$$

For the choice

$$(3.5) \quad k_n = \lfloor \text{const} \cdot n^{2/(d+2)} \rfloor,$$

the rate (3.4) is of order

$$(3.6) \quad O\left(\frac{\ln n}{n^{1/(d+2)}}\right).$$

Under mild condition, the following theorem states strong consistency and the rate of convergence of the excess length $[\Delta_n(x) - \Delta^*(x)]^+$:

Theorem 3.2. *Assume that μ has a density and $k_n/(\ln n)^2 \rightarrow \infty$ and $(\ln n)^d k_n/n \rightarrow 0$.*

(i) *If either the conditional quantiles at p^* and at $p^* + 1 - \alpha$ are unique, or the function $F(y | \cdot)$ satisfies the modified Lipschitz condition, then*

$$(3.7) \quad \lim_{n \rightarrow \infty} [\Delta_n(x) - \Delta^*(x)]^+ = 0$$

a.s. for μ -almost-all x .

(ii) *Under the modified Lipschitz condition we have that*

$$(3.8) \quad \mathbb{P}\{\Delta_n(x) - \Delta^*(x) \geq 0\} \leq 2n^{-2 \ln n} + 2e^{-3k_n/14 \mathbb{I}_{\ln n \geq 2C^*}}.$$

We get that $\mathbb{I}_{\ln n \geq 2C^*} = 1$, if n is large enough. Thus, the choice (3.5) and (3.8) imply the super-polynomial rate of convergence:

$$\mathbb{P}\{\Delta_n(x) - \Delta^*(x) \geq 0\} \leq 2n^{-2 \ln n} + 2e^{-3k_n/14} = O(1/n^{2 \ln n}).$$

It is an open problem whether (3.7) is satisfied without assuming anything on the conditional distribution function, or whether there is an estimate of the optimal prediction interval with distribution-free consistency of the excess length of the prediction interval.

4. Estimate the confidence set for classification

The previous results have some consequences in classification. For the classification, the label Y takes values in the finite set $\{1, \dots, K\}$ and the Bayes decision g^* minimizes the error probability:

$$g^*(x) = \arg \max_j P_j(x),$$

where

$$P_j(x) = \mathbb{P}\{Y = j | X = x\}, \quad j = 1, \dots, K.$$

A possibility for improving the confidence of the classification is the Bayes decision with rejection option, which means that for small values of $\max_j P_j(x)$ one does not make any decision, see Problem 2.5 in [7].

Another possibility is, that instead of deciding on a single label, we give a list of labels. Such list is called confidence set for classification. The problem is to construct a confidence set $C_n(x) = C_n(\mathcal{D}_n, x)$ such that for the *conditional coverage probability*

$$\mathbb{P}\{Y \in C(X) \mid X = x\} \geq 1 - \alpha,$$

where $0 < \alpha < 1$ is fixed, and the *cardinality* $|C(x)|$ of the set $C(x)$ is as small as possible. (Cf. [4] and [5].)

If the a posteriori probabilities $P_j(x)$, $j = 1, \dots, K$ are known, then the optimal confidence set can be constructed. Let $P_{i_1(x)}(x) \geq \dots \geq P_{i_K(x)}(x)$ be the ordered values of $P_1(x), \dots, P_K(x)$. Define $L^*(x)$ by the inequalities

$$\sum_{j=1}^{L^*(x)} P_{i_j(x)}(x) \geq 1 - \alpha > \sum_{j=1}^{L^*(x)-1} P_{i_j(x)}(x).$$

Then, the optimal confidence set is

$$C^*(x) = \{i_1(x), \dots, i_{L^*(x)}(x)\}.$$

If the a posteriori probabilities $P_j(x)$ are unknown, then an approximation of P_j is denoted by \tilde{P}_j , $j = 1, \dots, K$, from which the corresponding plug-in confidence set is as follows: Let $\tilde{P}_{i_1(x)}(x) \geq \dots \geq \tilde{P}_{i_K(x)}(x)$ be the ordered values of $\tilde{P}_1(x), \dots, \tilde{P}_K(x)$. Define $\tilde{L}(x)$ by the inequalities

$$\sum_{j=1}^{\tilde{L}(x)} \tilde{P}_{i_j(x)}(x) \geq 1 - \alpha - \tilde{t} > \sum_{j=1}^{\tilde{L}(x)-1} \tilde{P}_{i_j(x)}(x)$$

with a $\tilde{t} > 0$. Then, the plug-in confidence set is

$$\tilde{C}(x) = \{\tilde{i}_1(x), \dots, \tilde{i}_{\tilde{L}(x)}(x)\}.$$

In this way we underestimate the optimal confidence set.

Choose an integer k less than n , then the k -nearest-neighbor estimate of P_j is

$$(4.1) \quad P_{n,j}(x) = \frac{1}{k} \sum_{i=1}^k \mathbb{I}_{\{Y_{(n,i)}(x)=j\}}.$$

As before, from these estimates we derive the plug-in confidence set $C_n(x)$ such that C_n and $P_{n,j}$ correspond to \tilde{C} and \tilde{P}_j , respectively. Furthermore,

$$\tilde{t} = t_{k,n} := \ln n(1/\sqrt{k} + (k/n)^{1/d}).$$

Theorem 4.1. Assume that μ has a density. If $k = k_n$ such that $k_n/(\ln n)^2 \rightarrow \infty$ and $(\ln n)^d k_n/n \rightarrow 0$, then for any distribution of (X, Y) ,

$$(4.2) \quad \liminf_{n \rightarrow \infty} \mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} \geq 1 - \alpha$$

a.s. for μ -almost-all x . If, in addition, the modified Lipschitz condition is satisfied, then for any fixed x ,

$$(4.3) \quad \begin{aligned} & \mathbb{E} \{((1 - \alpha) - \mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\})^+\} \\ & = O\left(\ln n / \sqrt{k_n}\right) + O\left(\ln n (k_n/n)^{1/d}\right). \end{aligned}$$

Theorem 4.2. Assume that μ has a density and $k_n/(\ln n)^2 \rightarrow \infty$ and $(\ln n)^d k_n/n \rightarrow 0$. Under the modified Lipschitz condition we have that

$$(4.4) \quad \mathbb{P}\{|C_n(x)| - |C^*(x)| > 0\} \leq 2n^{-2 \ln n} + 2e^{-3k_n/14} \mathbb{I}_{\ln n \geq 2C^*}.$$

5. Proofs

Proof Theorem 2.1. We have that

$$\mathbb{P}\{Y_{n+1} \in C_n \mid Y_1, \dots, Y_n\} = F(f_{n,hi}) - F(f_{n,lo}).$$

and

$$\begin{aligned} & |[F(f_{n,hi}) - F(f_{n,lo})] - (1 - \alpha)| \\ & \leq |[F(f_{n,hi}) - F(f_{n,lo})] - (1 - \alpha - 2 \ln n / \sqrt{n})| + 2 \ln n / \sqrt{n} \\ & = |[F(f_{n,hi}) - F(f_{n,lo})] - [F_n(f_{n,hi}) - F_n(f_{n,lo})]| + 2 \ln n / \sqrt{n} \\ & \leq 2 \sup_y |F(y) - F_n(y)| + 2 \ln n / \sqrt{n}. \end{aligned}$$

Thus, Theorem 12.4 in Devroye, Györfi and Lugosi [7] implies that

$$\begin{aligned} & \mathbb{P}\{|[F(f_{n,hi}) - F(f_{n,lo})] - (1 - \alpha)| \geq \varepsilon\} \\ & \leq \mathbb{P}\{2 \sup_y |F(y) - F_n(y)| \geq \varepsilon/2\} + \mathbb{I}_{2 \ln n / \sqrt{n} \geq \varepsilon/2} \\ & \leq 8(n + 1)e^{-n\varepsilon^2/128} + \mathbb{I}_{\ln n / \sqrt{n} \geq \varepsilon/4}. \end{aligned}$$

Furthermore,

$$(5.1) \quad \begin{aligned} \Delta_n - \Delta^* & = D_n(p_n) - D(p^*) \\ & \leq D_n(p^*) - D(p^*) \\ & \leq (F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n / \sqrt{n}))^+ \\ & \quad + (F_{n,low}^{-1}(p^* + 1 - \alpha - \ln n / \sqrt{n}) - F_{low}^{-1}(p^* + 1 - \alpha))^+. \end{aligned}$$

One has

$$\begin{aligned} & \mathbb{P}\{F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n/\sqrt{n}) \geq 0\} \\ & \leq \mathbb{P}\{F_n(F_{up}^{-1}(p^*)) \geq p^* + \ln n/\sqrt{n}\} \\ & = \mathbb{P}\{F_n(F_{up}^{-1}(p^*)) - F(F_{up}^{-1}(p^*)) \geq \ln n/\sqrt{n}\}. \end{aligned}$$

The Hoeffding inequality implies that

$$\mathbb{P}\{F_{up}^{-1}(p^*) - F_{n,up}^{-1}(p^* + \ln n/\sqrt{n}) \geq 0\} \leq e^{-2n(\ln n/\sqrt{n})^2} = 1/n^{2 \ln n}.$$

Similarly, we get that

$$\mathbb{P}\{F_{n,low}^{-1}(p^* + 1 - \alpha - \ln n/\sqrt{n}) - F_{low}^{-1}(p^* + 1 - \alpha) \geq 0\} \leq 1/n^{2 \ln n}.$$

□

Proof of Theorem 3.1. One has that

$$\mathbb{P}\{Y_{n+1} \in C_{k,n}(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} = F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x).$$

Devroye [6] proved that $k_n/\ln n \rightarrow \infty$ and $k_n/n \rightarrow 0$ imply that

$$(5.2) \quad |F(y \mid x) - F_{k_n,n}(y \mid x)| \rightarrow 0$$

a.s. for all y and for μ -almost all x , from which we get

$$(5.3) \quad \sup_y |F(y \mid x) - F_{k_n,n}(y \mid x)| \rightarrow 0$$

a.s. for μ -almost all x , where we used the argument of the standard proof of Glivenko-Cantelli theorem showing that the pointwise convergence of empirical distribution functions implies the uniform convergence. We have that

$$\begin{aligned} & |[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] - (1 - \alpha)| \\ & \leq |[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] - (1 - \alpha - 2t_{k,n})| + 2t_{k,n} \\ & = |[F(f_{k,n,hi}(x) \mid x) - F(f_{k,n,lo}(x) \mid x)] \\ & \quad - [F_{k,n}(f_{k,n,hi}(x) \mid x) - F_{k,n}(f_{k,n,lo}(x) \mid x)]| + 2t_{k,n} \\ (5.4) \quad & \leq 2 \sup_y |F(y \mid x) - F_{k,n}(y \mid x)| + 2t_{k,n}. \end{aligned}$$

Under the conditions of the theorem

$$t_{k_n,n} \rightarrow 0.$$

Therefore (5.3) and (5.4) imply (3.3). Because of (5.4), (3.4) is proved if

$$(5.5) \quad \mathbb{E} \left\{ \sup_y |F(y \mid x) - F_{k,n}(y \mid x)| \right\} = O(\ln k/\sqrt{k}) + O((k/n)^{1/d}).$$

Put

$$\bar{F}_{k,n}(y | x) = \mathbb{E}\{F_{k,n}(y | x) | X_1, \dots, X_n\} = \frac{1}{k} \sum_{i=1}^k F(y | X_{(n,i)}(x)).$$

Concerning (5.5), we show that

$$(5.6) \quad \mathbb{E} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| \right\} = O(\ln k / \sqrt{k})$$

and

$$(5.7) \quad \mathbb{E} \left\{ \sup_y |\bar{F}_{k,n}(y | x) - F(y | x)| \right\} = O((k/n)^{1/d}).$$

Proof of (5.6): We have

$$\begin{aligned} & \mathbb{E} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| \right\} \\ &= \mathbb{E} \left\{ \mathbb{E} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| | X_1, \dots, X_n \right\} \right\} \\ &= \mathbb{E} \left\{ \int_0^1 \mathbb{P} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| \geq s | X_1, \dots, X_n \right\} ds \right\}. \end{aligned}$$

Conditioning with respect to X_1, \dots, X_n we note that $F_{k,n}(y | x)$ is an arithmetic mean of independent $\{0, 1\}$ valued random variables. Although these random variables are not identically distributed, we can extend the proof of Theorem 12.4 in Devroye, Györfi and Lugosi [7], which is a sharpened version of the Glivenko-Cantelli theorem. Thus, we obtain that

$$\mathbb{P} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| \geq s | X_1, \dots, X_n \right\} \leq 2(k+1)e^{-ks^2/32},$$

$0 < s < 1$, which implies

$$\begin{aligned} \mathbb{E} \left\{ \sup_y |F_{k,n}(y | x) - \bar{F}_{k,n}(y | x)| \right\} &\leq \int_0^1 \min\{1, 2(k+1)e^{-ks^2/32}\} ds \\ &\leq \varepsilon + 2(k+1) \int_\varepsilon^1 e^{-ks^2/32} ds \\ &\leq \varepsilon + 2(k+1) \frac{2\sqrt{2}}{\sqrt{k}} \int_{k\varepsilon^2/32}^\infty e^{-t} t^{-1/2} dt \\ &\leq \varepsilon + \frac{64}{\varepsilon} e^{-k\varepsilon^2/32}, \end{aligned}$$

which yields (5.6) by the choice $\varepsilon = \ln k / \sqrt{k}$.

Proof of (5.7): The modified Lipschitz condition (3.2) implies that for each y ,

$$\begin{aligned} |\bar{F}_{k,n}(y | x) - F(y | x)| &\leq \frac{1}{k} \sum_{i=1}^k |F(y | X_{(n,i)}(x)) - F(y | x)| \\ &\leq C^* \frac{1}{k} \sum_{i=1}^k \mu(S_{x, \|x - X_{(n,i)}(x)\|})^{1/d} \\ &\leq C^* \mu(S_{x, \|x - X_{(n,k)}(x)\|})^{1/d}. \end{aligned}$$

For i.i.d. uniformly distributed U_1, \dots, U_n , let $U_{(1,n)}, \dots, U_{(n,n)}$ denote the corresponding order statistic. If μ has a density, then from Section 1.2 in Biau and Devroye [2] we have that

$$(5.8) \quad \mu(S_{x, \|x - X_{(n,k)}(x)\|}) \stackrel{D}{=} U_{(k,n)}.$$

Thus, for any fixed y and $s > 0$ with $(s/C^*)^d > k/n$, the Bernstein inequality implies that

$$\begin{aligned} &\mathbb{P} \left\{ \sup_y |\bar{F}_{k,n}(y | x) - F(y | x)| \geq s \right\} \\ &\leq \mathbb{P} \left\{ C^* U_{(k,n)}^{1/d} \geq s \right\} \\ &= \mathbb{P} \left\{ U_{(k,n)} \geq (s/C^*)^d \right\} \\ &= \mathbb{P} \left\{ \frac{1}{n} \sum_{i=1}^n (\mathbb{I}_{U_i \leq (s/C^*)^d} - (s/C^*)^d) < -(s/C^*)^d - k/n \right\} \\ &\leq e^{-\frac{n[(s/C^*)^d - k/n]^2}{2(s/C^*)^d + 2[(s/C^*)^d - k/n]/3}}. \end{aligned}$$

Therefore

$$\begin{aligned} &\mathbb{E} \left\{ \sup_y |\bar{F}_{k,n}(y | x) - F(y | x)| \right\} \\ &= \int_0^1 \mathbb{P} \left\{ \sup_y |\bar{F}_{k,n}(y | x) - F(y | x)| \geq s \right\} ds \\ &\leq 2C^*(k/n)^{1/d} + \int_{2C^*(k/n)^{1/d}}^\infty e^{-\frac{n[(s/C^*)^d - k/n]^2}{2(s/C^*)^d + 2[(s/C^*)^d - k/n]/3}} ds \\ &\leq 2C^*(k/n)^{1/d} + \int_{2C^*(k/n)^{1/d}}^\infty e^{-3n(s/C^*)^d/32} ds \\ &\leq 2C^*(k/n)^{1/d} + O((1/n)^{1/d}) \\ &= O((k/n)^{1/d}), \end{aligned}$$

and so we get (5.7). □

Proof of Theorem 3.2. Notice that

$$\begin{aligned}
 \Delta_n(x) - \Delta^*(x) &= D_n(p_n) - D(p^*) \\
 &\leq D_n(p^*) - D(p^*) \\
 &\leq (F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}))^+ \\
 (5.9) \quad &+ (F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha))^+.
 \end{aligned}$$

Proof of (ii): We have that

$$\begin{aligned}
 &\mathbb{P}\left\{F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}) \geq 0\right\} \\
 &\leq \mathbb{P}\left\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq p^* + t_{k,n}\right\} \\
 &= \mathbb{P}\left\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(1/\sqrt{k} + (k/n)^{1/d})\right\} \\
 &\leq \mathbb{P}\left\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - \bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n/\sqrt{k}\right\} \\
 &+ \mathbb{P}\left\{\bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(k/n)^{1/d}\right\}.
 \end{aligned}$$

As in the previous proofs, the Hoeffding inequality and the Bernstein inequality imply that

$$\begin{aligned}
 \mathbb{P}\left\{F_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - \bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n/\sqrt{k}\right\} &\leq e^{-2k(\ln n/\sqrt{k})^2} \\
 &= n^{-2\ln n},
 \end{aligned}$$

and for $\ln n \geq 2C^*$,

$$\begin{aligned}
 &\mathbb{P}\left\{\bar{F}_{k,n}(F_{x,up}^{-1}(p^*) \mid x) - F(F_{x,up}^{-1}(p^*) \mid x) \geq \ln n(k/n)^{1/d}\right\} \\
 &\leq \mathbb{P}\left\{C^*U_{(k,n)}^{1/d} \geq \ln n(k/n)^{1/d}\right\} \\
 &\leq \mathbb{P}\left\{U_{(k,n)} \geq 2k/n\right\} \\
 &= \mathbb{P}\left\{\frac{1}{n} \sum_{i=1}^n (\mathbb{I}_{U_i \leq 2k/n} - \mathbb{E}\{\mathbb{I}_{U_i \leq 2k/n}\}) \leq -k/n\right\} \\
 &\leq e^{-\frac{n(k/n)^2}{4(k/n)+2(k/n)/3}} \\
 &= e^{-3k/14}.
 \end{aligned}$$

Thus,

$$\mathbb{P}\left\{F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}) \geq 0\right\} \leq n^{-2\ln n} + e^{-3k/14} \mathbb{I}_{\ln n \geq 2C^*}.$$

The proof of

$$\begin{aligned} & \mathbb{P} \left\{ F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha) \geq 0 \right\} \\ & \leq n^{-2 \ln n} + e^{-3k/14} \mathbb{I}_{\ln n \geq 2C^*} \end{aligned}$$

is similar.

Proof of (i): Because of (5.9) we have to prove that

$$(F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}))^+ \rightarrow 0$$

and

$$(F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha))^+ \rightarrow 0$$

a.s. for μ -almost-all x .

If the conditional quantiles at p^* and at $p^* + 1 - \alpha$ are unique, then Theorem 1 in Hansmann and Kohler [10] and (5.2) imply

$$\begin{aligned} (F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^* + t_{k,n}))^+ & \leq (F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^*))^+ \\ & \leq |F_{x,up}^{-1}(p^*) - F_{x,k,n,up}^{-1}(p^*)| \\ & \rightarrow 0 \end{aligned}$$

and

$$\begin{aligned} & (F_{x,k,n,low}^{-1}(p^* + 1 - \alpha - t_{k,n}) - F_{x,low}^{-1}(p^* + 1 - \alpha))^+ \\ & \leq (F_{x,k,n,low}^{-1}(p^* + 1 - \alpha) - F_{x,low}^{-1}(p^* + 1 - \alpha))^+ \\ & \leq |F_{x,k,n,low}^{-1}(p^* + 1 - \alpha) - F_{x,low}^{-1}(p^* + 1 - \alpha)| \\ & \rightarrow 0 \end{aligned}$$

a.s. for μ -almost-all x .

If the function $F(y | \cdot)$ satisfies the modified Lipschitz condition, and $\mathbb{I}_{\ln n \geq 2C^*} = 1$ for n large enough, then (3.8) implies that

$$\mathbb{P}\{\Delta_n(x) - \Delta^*(x) \geq 0\} \leq 2n^{-2 \ln n} + 2e^{-3k_n/14}$$

and so the condition $k_n / \ln n \rightarrow \infty$ together with the Borel-Cantelli lemma implies the strong consistency (3.7). \square

Proof of Theorem 4.1. For the plug-in confidence set, we can bound the conditional coverage probability and the excess size as follows.

$$\begin{aligned} \mathbb{P}\{Y \in \tilde{C}(X) \mid X = x\} & = \sum_{j \in \tilde{C}(x)} P_j(x) \\ & \geq \sum_{j \in \tilde{C}(x)} \tilde{P}_j(x) - \sum_{j=1}^K |P_j(x) - \tilde{P}_j(x)| \\ (5.10) \quad & \geq 1 - \alpha - \tilde{t} - \sum_{j=1}^K |P_j(x) - \tilde{P}_j(x)|. \end{aligned}$$

Furthermore,

$$\begin{aligned}
 \mathbb{I}_{|\tilde{C}(x)| - |C^*(x)| > 0} &= \mathbb{I}_{\tilde{L}(x) > L^*(x)} \\
 &= \mathbb{I}_{\tilde{L}(x) - 1 \geq L^*(x)} \\
 &\leq \mathbb{I}_{1 - \alpha - \tilde{t} > \sum_{j=1}^{L^*(x)} \tilde{P}_{i_j(x)}(x)} \\
 &\leq \mathbb{I}_{1 - \alpha - \tilde{t} > \sum_{j=1}^{L^*(x)} P_{i_j(x)}(x) - \sum_{j=1}^K |P_j(x) - \tilde{P}_j(x)|} \\
 &\leq \mathbb{I}_{1 - \alpha - \tilde{t} > 1 - \alpha - \sum_{j=1}^K |P_j(x) - \tilde{P}_j(x)|} \\
 (5.11) \quad &= \mathbb{I}_{\tilde{t} < \sum_{j=1}^K |P_j(x) - \tilde{P}_j(x)|}.
 \end{aligned}$$

(5.10) implies that

$$\begin{aligned}
 (5.12) \quad &\mathbb{P}\{Y_{n+1} \in C_n(X_{n+1}) \mid \mathcal{D}_n, X_{n+1} = x\} \\
 &\geq 1 - \alpha - t_{k,n} - \sum_{j=1}^K |P_j(x) - P_{n,j}(x)|.
 \end{aligned}$$

Under the conditions of the theorem $t_{k,n} \rightarrow 0$. Devroye [6] proved that $k_n / \ln n \rightarrow \infty$ and $k_n/n \rightarrow 0$ imply that

$$(5.13) \quad |P_j(x) - P_{n,j}(x)| \rightarrow 0$$

a.s. for all j and for μ -almost all x , from which we get the first half of the theorem. Because of (5.12), (4.3) is proved if

$$(5.14) \quad \mathbb{E}\{|P_j(x) - P_{n,j}(x)|\} = O\left(\ln k / \sqrt{k}\right) + O\left((k/n)^{1/d}\right).$$

This last step can be verified in the same way as in the proof of Theorem 3.1. \square

Proof of Theorem 4.2. (5.11) implies that

$$\begin{aligned}
 \mathbb{P}\{|C_n(x)| - |C^*(x)| > 0\} &\leq \mathbb{P}\left\{t_{k,n} < \sum_{j=1}^K |P_j(x) - P_{n,j}(x)|\right\} \\
 &\leq 2n^{-2 \ln n} + 2e^{-3k_n/14} \mathbb{I}_{\ln n \geq 2C^*},
 \end{aligned}$$

where the last step can be made as in the proof of Theorem 3.2. \square

References

- [1] Barber, R. F., Candés, E. J., Ramdas, A. and Tibshirani, R. J.: The limits of distribution-free conditional predictive inference. (submitted), 2020.
- [2] Biau, G. and Devroye, L.: *Lectures on the Nearest Neighbor Method*, Springer-Verlag, New York, 2015.

- [3] Chaudhuri K. and Dasgupta, S.: Rates of convergence for nearest neighbor classification. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence and K. Q. Weinberger, editors, 3437–3445, 2014.
- [4] Chzhen, E., Denis, C. and Hebiri, M.: Minimax semi-supervised set-valued approach to multi-class classification. (submitted), 2020.
- [5] Denis, C. and Hebiri, M.: Confidence sets with expected sizes for multiclass classification. *Journal of Machine Learning Research*, 18:1–28, 2017.
- [6] Devroye, L.: Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61:467–481, 1982.
- [7] Devroye, L., Györfi, L. and Lugosi, G.: *A Probabilistic Theory of Pattern Recognition*, Springer-Verlag, New York, 1996.
- [8] Döring, M., Györfi, L. and Walk, H.: Exact rate of convergence of k -nearest-neighbor classification rule, *Journal of Machine Learning Research*, 18:1–16, 2018.
- [9] Györfi, L., Kohler, M., Krzyżak, A. and Walk, H.: *A Distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New York, 2002.
- [10] Hansmann, M. and Kohler, M.: Estimation of conditional quantiles from data with additional measurement errors, *J. Statistical Planning and Inference*, 200:176–195, 2019.
- [11] Lei, J. and Wasserman, L.: Distribution-free prediction bands for nonparametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76:71–96, 2014.
- [12] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J. and Wasserman, L.: Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113:1094–1111, 2018.
- [13] Lei, J., Robins, J. and Wasserman, L.: Distribution-free prediction sets. *Journal of the American Statistical Association*, 108:278–287, 2013.
- [14] Papadopoulos, H., Proedrou, K., Vovk, V. and Gammernan, A.: Inductive confidence machines for regression. In *European Conference on Machine Learning*, pages 345–356. Springer-Verlag, New York, 2002.
- [15] Romano, Y., Patterson, E. and Candés, E. J.: Conformalized quantile regression, *Advances in Neural Information Processing Systems 32 (NIPS 2019)*, pp. 3538–3548, 2019.
- [16] Vovk, V.: Conditional validity of inductive conformal predictors. *Machine Learning*, 92:349–376, 2013.

Budapest University of Technology and Economics
 Department of Computer Science and Information Theory
 Magyar Tudósok krt. 2., Budapest, H-1117, Hungary
 e-mail: gyorfi@cs.bme.hu

University of Stuttgart
 Institute of Stochastics and Applications
 Pfaffenwaldring 57, D-70569 Stuttgart, Germany
 e-mail: harro.walk@mathematik.uni-stuttgart.de