On Nonparametric Regression using Data Depth

Joydeep Chowdhury and Probal Chaudhuri
Indian Statistical Institute, Kolkata

Abstract

We investigate nonparametric regression methods based on statistical depth functions. These nonparametric regression procedures can be used in situations, where the response is multivariate and the covariate is a random element in a metric space. This includes regression with functional covariate as a special case. Our objective is to study different features of the conditional distribution of the response given the covariate. We construct measures of the center and the spread of the conditional distribution using depth based nonparametric regression procedures. We establish the asymptotic consistency of those measures and develop a test for heteroscedasticity based on the measure of conditional spread. The usefulness of the methodology is demonstrated in some real datasets. In one dataset consisting of Italian household expenditure data for the period 1973 to 1992, we regress the expenditure for different items on their prices. In another dataset, our responses are the nutritional contents of different meat samples measured by their protein, fat and moisture contents, and the functional covariate is the absorbance spectra of the meat samples.

Keywords: Central region, Conditional depth, Conditional median, Conditional spread, Conditional trimmed mean, Heteroscedasticity.

1 Introduction

A statistical depth function provides an ordering of the points such that the points close to the centre of the distribution have higher depth values than those that are away from the centre. Several depth functions are defined in the literature, e.g., the halfspace depth (Tukey, 1975; Donoho and Gasko, 1992), the simplicial depth (Liu, 1990), the spatial depth (Vardi and Zhang, 2000; Serfling, 2002), the projection depth (Zuo and Serfling, 2000a,b), and many others. Depth functions have been used earlier for various purposes like detecting outliers (Chen et al., 2009), clustering (Jörnsten, 2004) and classification (Ghosh and Chaudhuri, 2005; Dutta and Ghosh, 2012; Li et al., 2012). We develop nonparametric regression methods using statistical depth functions. The response in our setup is multivariate, while the covariate is a random element in a separable metric space, which includes the cases of finite dimensional as well as infinite dimensional or function-valued covariates. The traditional nonparametric regression methods estimate the conditional mean or median (see, e.g., Härdle (1990), Fan and Gijbels (1996), Wand and Jones (1994), etc.), and provide information about only the center of the conditional distribution of the response given the covariate. These methods are not useful if one is interested in other features of the conditional distribution, like the conditional spread and heteroscedasticity in the sample. The methods we develop will provide simultaneous
information about the center as well as other parts of the conditional distribution, and help us construct measures of conditional spread of the response.

Our methods based on statistical depth are inspired from Tukey’s boxplot (Tukey, 1975, 1977). The boxplot for a univariate random variable provides information about the center of the distribution through the median as well as the spread and the skewness through the first and third quartiles. Let us consider the following two examples, which demonstrate the inadequacy of regression procedures that consider only the center of the conditional distribution.

**Example 1:** In our first example, we consider the BudgetItaly Data, which is available in the Ecdat package in R. This dataset contains the following information on budget shares of 1729 Italian households for the period 1973 to 1992: the shares of food expenditure, housing and fuels expenditure and miscellaneous expenditure in the total budget along with the corresponding food price, housing and fuels price and miscellaneous price. We consider the variable pair of food share and housing and fuels share as the response, and the variable pair of food price and housing and fuels price as the covariate.

To carry out nonparametric regression, we construct a neighborhood consisting of 50 nearest covariate values for each covariate value, based on the euclidean metric. The collection of the corresponding response values for a covariate value represents the conditional distribution of the response given that covariate value. We call these response values the local response values for the covariate. We choose 4 covariate values, and plot their local response values and the marginal boxplots for the local response values in Figure 1. The selected covariate values are mentioned over the corresponding columns in the plot. The local boxplots, which are the boxplots for the first and the second coordinates of the local response values, estimate the conditional boxplots of the two coordinates of the response given the covariate. They clearly reflect the varying conditional spread and skewness of the response over the covariate values, which could not be detected from mean or median regression. But as indicated by the scatter plots of the local response values, the two coordinates are correlated. Consequently, it is desirable to investigate the features of the joint distribution of the two coordinates given the covariate, which cannot be done using the local marginal boxplots.

**Example 2:** Our second example deals with the Tecator Data, which is available in the caret package in R. This dataset contains the percentage values of moisture, fat and protein contents of 215 meat samples along with their absorbance spectra in the wavelength range 850–1050 nm. The moisture, the fat and the protein contents were measured by analytical chemistry, while a Tecator Infratec Food and Feed Analyzer was used to record the absorbance spectrum. Being able to predict the nutritional contents of a meat sample from its absorbance spectra is economically beneficial since obtaining the spectra is relatively cheaper. We consider the pair of fat and protein contents as the response, and the curve of absorbance spectra as the covariate. So, the response is bivariate and the covariate is functional. The covariate is considered to be a random element in the $L_2$ space. We do the same analysis as in **Example 1**, with the euclidean metric replaced by the $L_2$ metric and changing the nearest neighbor size to 30. We plot the results, along with the chosen covariate curves, in Figure 2. Here again, the local boxplots reflect the varying conditional spread and skewness of the response over the covariate curves, and the scatter plots reflect the dependence between the two coordinates of the local response values. This again implies the importance of investigating the features of the bivariate conditional distribution of the response given the covariate curves.

The preceding two examples demonstrate the usefulness of the boxplots in yielding
Figure 1: The local response values (1st row), the local boxplots for the 1st (2nd row) and the 2nd coordinates (3rd row) of the response variable in the BudgetItaly Data.
Figure 2: The selected covariate curves (1st row), the local response values (2nd row), the local boxplots for the 1st (3rd row) and the 2nd coordinates (4th row) of the response variable in the Tecator Data.
information about univariate data, as well as their shortcomings when the data is not
univariate. We develop depth based regression methods, which is applicable for multi-
variate responses. Earlier attempts to generalize boxplot using statistical depth functions
in unconditional setup can be found in [Rousseeuw et al. (1999) and Liu et al. (1999) for
bivariate data and in Sun and Genton (2011) for functional data. For a univariate distri-
bution, the centre of the boxplot, which is the median, is the point having the largest
depth value, and the box contains 50% of the observations which have higher depth values
than the points outside the box. These basic features of the boxplot are utilized to gen-
eralize boxplots for multivariate conditional distributions in the following section. Such
a boxplot based regression has fundamental connections with the well-known quantile
regression [Koenker 2005].

2 Conditional depth and central regions

Let \( Y \) be a random vector in \( \mathbb{R}^p \) and \( X \) be a random element in a complete separable
metric space \((C, d)\). Let \( \mu(\cdot | z) \) be the conditional probability measure of \( Y \) given \( X = z \),
and \( \rho(\cdot | z) \) be a conditional depth function on the response space \( \mathbb{R}^p \) related to \( \mu(\cdot | z) \).
Values of the depth functions usually lie between 0 and 1 and consequently it is uniformly
bounded. Let \( x \) be a fixed element in \( C \). Define \( D(\alpha | x) = \{ y \in \mathbb{R}^p \mid \rho(y | x) \geq \alpha \} \)
for \( \alpha \in \mathbb{R} \). \( D(\alpha | x) \) is called the conditional \( \alpha \)-trimmed region of \( Y \) given \( X = x \)
corresponding to the conditional depth \( \rho(\cdot | x) \) (cf. Zuo and Serfling (2000a,b)). For
\( 0 \leq r < 1 \), let \( \alpha(r) = \sup\{ \alpha \mid \mu(\{ y \mid \rho(y | x) \geq \alpha \}) \geq r \} \), which is finite when \( \rho(y | x) \) is uniformly
bounded over \( y \). We define the set \( D(\alpha(r) | x) \) as the conditional 100\( \% \) central region of \( Y \)
given \( X = x \) with respect to the conditional depth \( \rho(\cdot | x) \). Clearly, \( \mu(D(\alpha(r) | x) | x) \geq r \),
and \( \mu(D(\alpha(r) | x) | x) = r \) whenever \( \mu(\{ y \in \mathbb{R}^p \mid \rho(y | x) = \alpha(r) \}) | x) = 0 \). Further, for
any \( y_1 \in D(\alpha(r) | x) \) and \( y_2 \notin D(\alpha(r) | x) \), \( \rho(y_1 | x) \geq \rho(y_2 | x) \). When the response is
univariate, the conditional 50\( \% \) central region corresponds to the box in the conditional
boxplot of the response.

To estimate \( \rho(\cdot | x) \), \( D(\alpha | x) \) and \( D(\alpha(r) | x) \) based on a random sample \((X_1, Y_1), \ldots, (X_n, Y_n)\),
we adopt a nonparametric regression procedure. Let \( W_{i,n}(x) \) be the weight
on the observation pair \((X_i, Y_i), i = 1, \ldots, n, \) \( W_{i,n}(x) \geq 0 \) for each \( i \) and \( \sum_{i=1}^{n} W_{i,n}(x) = 1 \).
The sample conditional probability measure of \( Y \) given \( X = x \) is defined as

\[
\mu_n(B | x) = \sum_{i=1}^{n} I(Y_i \in B)W_{i,n}(x),
\]

where \( B \) is any Borel set. The conditional sample depth function \( \rho_n(\cdot | x) \) is defined based on
\( \mu_n(\cdot | x) \) in the same way as the conditional population depth function \( \rho(\cdot | x) \) is defined
based on \( \mu(\cdot | x) \). The conditional sample \( \alpha \)-trimmed region of \( Y \) given \( X = x \) is defined as
\( D_n(\alpha | x) = \{ y \in \mathbb{R}^p \mid \rho_n(y | x) \geq \alpha \} \). The conditional sample 100\( \% \) central region of \( Y \)
given \( X = x \) is \( D_n(\alpha_n(r) | x) \), where \( \alpha_n(r) = \sup\{ \alpha \mid \mu_n(D_n(\alpha | x) | x) \geq r \} \). In
an unconditional setup, a version of the central region was defined in [Liu et al. (1999)
p. 788], and [Chowdhury and Chaudhuri (2016)] defined conditional maximal depth sets
based on the spatial depth. However, neither [Liu et al. (1999)] nor [Chowdhury and
Chaudhuri (2016)] investigated theoretical properties like asymptotic consistency of the
sample central regions and maximal depth sets.

We now describe the conditional versions of several well-known depth functions. The conditional
halfspace depth [Tukey (1975)] and [Donoho and Gasko (1992)] of \( Y \) given \( X = x \) is
The estimate \( \rho(Y) \) of \( Y \) is defined as \( \rho(Y) = \inf \{ \mu(\{ v \in \mathbb{R}^p \mid u'v \geq u'y \} \mid x) \mid u \in \mathbb{R}^p \} \). Its estimate is \( \rho_n(y \mid x) = \inf \{ \mu_n(\{ v \in \mathbb{R}^p \mid u'v \geq u'y \} \mid x) \mid u \in \mathbb{R}^p \} \). The conditional Spatial Depth (Vardi and Zhang 2000; Serfling 2002) is defined as \( \rho(y \mid x) = 1 - \| E[|y - Y|^{-1}(y - Y) \mid X = x] \| \).

The estimate \( \rho_n(y \mid x) \) is given by

\[
\rho_n(y \mid x) = 1 - \left( \sum_{i=1}^{n} \| y - Y_i \|^{-1} (y - Y_i) W_{i,n}(x) \right).
\]

The conditional Projection Depth (Zuo and Serfling 2000a,b) is defined as

\[
\rho(y \mid x) = \left[ 1 + \sup_{\| u \|=1} \frac{|u'y - M(u'y \mid x)|}{M(|u'y - M(u'y \mid x)| \mid x)} \right]^{-1},
\]

where \( M(u'y \mid x) \) and \( M(|u'y - M(u'y \mid x)| \mid x) \) are the conditional medians of \( u'y \) and \( |u'y - M(u'y \mid x)| \) given \( X = x \), respectively. The conditional sample projection depth is

\[
\rho_n(y \mid x) = \left[ 1 + \sup_{\| u \|=1} \frac{|u'y - M_n(u'y \mid x)|}{M_n(|u'y - M_n(u'y \mid x)| \mid x)} \right]^{-1},
\]

where \( M_n(u'y \mid x) \) and \( M_n(|u'y - M_n(u'y \mid x)| \mid x) \) are the sample analogues of \( M(u'y \mid x) \) and \( M(|u'y - M(u'y \mid x)| \mid x) \), respectively. Let \( Y^{(1)}, \ldots, Y^{(p+1)} \) be independent copies of \( Y \in \mathbb{R}^p \), and let \( \Delta(Y^{(1)}, \ldots, Y^{(p+1)}) \) denote the \( p \)-dimensional simplex formed from \( Y^{(1)}, \ldots, Y^{(p+1)} \). The conditional simplicial depth (Liu 1990) is defined as

\[
\rho(y \mid x) = P \left[ y \in \Delta(Y^{(1)}, \ldots, Y^{(p+1)}) \mid X = x \right].
\]

The weights \( \{ W_{i,n}(x) \} \) are constructed based on the covariate values \( X_1, \ldots, X_n \). There are several methods of selecting such weights. In the kernel regression method, we choose a kernel function \( K(\cdot) \) and a bandwidth \( h \), and the weight \( W_{i,n}(x) \) is

\[
W_{i,n}(x) = \frac{K(h^{-1}d(x, X_i))}{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))},
\]

where \( d(\cdot, \cdot) \) is the metric in the covariate space. This leads to a Nadaraya-Watson type kernel estimate (Nadaraya 1964; Watson 1964). In the nearest neighbor method, we choose a positive integer \( k \) for the number of nearest neighbors to be considered, and define

\[
h(x, k, n) = \min \left\{ h \left| \sum_{i=1}^{n} I(d(x, X_i) \leq h) \geq k \right. \right\}.
\]

The weight \( W_{i,n}(x) \) in this case is

\[
W_{i,n}(x) = \frac{I(d(x, X_i) \leq h(x, k, n))}{\sum_{i=1}^{n} I(d(x, X_i) \leq h(x, k, n))}.
\]
2.1 Median and trimmed mean regression

The depth-based conditional median \( \mathbf{m}(\mathbf{x}) \) of \( Y \) given \( \mathbf{X} = \mathbf{x} \) with respect to the conditional depth \( \rho(\cdot | \mathbf{x}) \) is a point such that \( \rho(\mathbf{m}(\mathbf{x}) | \mathbf{x}) \geq \rho(y | \mathbf{x}) \) for every \( y \). Note that \( \mathbf{m}(\mathbf{x}) \) may not be unique. For a univariate response and the usual depth functions like the conditional halfspace depth, the point \( \mathbf{m}(\mathbf{x}) \) becomes the usual conditional median, which can be viewed as the statistical center of the conditional boxplot, and the set \( D(\alpha(r) | \mathbf{x}) \) becomes the conditional interquartile interval for \( r = 0.5 \), which corresponds to the box in the conditional boxplot as already noted. \( \mathbf{m}(\mathbf{x}) \) along with the conditional central region \( D(\alpha(r) | \mathbf{x}) \) yields information about the centre and the spread of the conditional distribution of the response. The sample conditional median \( \mathbf{m}_n(\mathbf{x}) \) is a point such that \( \rho_n(\mathbf{m}_n(\mathbf{x}) | \mathbf{x}) \geq \rho_n(y | \mathbf{x}) \) for every \( y \).

Trimmed means based on depth functions in an unconditional setup were investigated earlier in [Donoho and Gasko (1992), Liu et al. (1999), Zuo (2006), Massé (2009), etc.]. The conditional 100\( r \)% trimmed mean \( \mathbf{m}(r | \mathbf{x}) \) of \( Y \) given \( \mathbf{X} = \mathbf{x} \) can be defined as \( \mathbf{m}(r | \mathbf{x}) = \{ \int \mathbf{y} I(\mathbf{y} \in D(\alpha(1-r) | \mathbf{x})) \mu(d\mathbf{y} | \mathbf{x}) / \mu(D(\alpha(1-r) | \mathbf{x}) | \mathbf{x}) \}. \) Unlike the conditional median, the conditional trimmed mean is always unique. For a real valued response, the depth-based conditional trimmed mean coincides with the usual conditional trimmed mean. The sample conditional 100\( r \)% trimmed mean \( \mathbf{m}_n(r | \mathbf{x}) \) is \( \mathbf{m}_n(r | \mathbf{x}) = \{ \int \mathbf{y} I(\mathbf{y} \in D_n(\alpha_n(1-r) | \mathbf{x})) \mu_n(d\mathbf{y} | \mathbf{x}) / \mu_n(D_n(\alpha_n(1-r) | \mathbf{x}) | \mathbf{x}) \}. \) Conditional trimmed means can be used to detect conditional skewness in the data. When the distribution is symmetric, the trimmed means lie away from the median.

2.2 Data demonstration

We demonstrate the conditional central regions, the conditional medians and the conditional trimmed means in the two real data examples considered in section 1. We choose the weights \( \{W_{i,n}(\mathbf{x})\} \) in a way which ensures the asymptotic consistency of the estimates. For this, we have employed the nearest neighbor approach, where the integer \( k = k_n \) is

\[
k_n = \lfloor (\log n)^2 \rfloor + 1, \tag{2.1}
\]

where \( \lfloor r \rfloor \) being the largest integer less than or equal to the real number \( r \). So,

\[
h(\mathbf{x}, k_n, n) = \min \left\{ h \left| \sum_{i=1}^{n} I(d(\mathbf{x}, \mathbf{X}_i) \leq h) > (\log n)^2 \right. \right\}.
\]

This approach is equivalent to the kernel method of choosing the weights when the kernel function is \( K(u) = I(0 \leq u \leq 1) \), and bandwidth \( h = h(\mathbf{x}, k_n, n) \). We shall see in section 4 that such a choice ensures the asymptotic consistency of the estimates. In the demonstrations, we use the conditional halfspace depth. With our choice of the weights, the conditional sample depth \( \rho_n(\cdot | \mathbf{x}) \) becomes the corresponding sample depth based on the local response values of \( \mathbf{x} \), which we compute using algorithms by [Rousseeuw and Ruts (1996) and Rousseeuw and Struyf (1998)].

We present the 50% conditional central regions for the conditional halfspace depth corresponding to four selected covariate values in the BudgetItaly Data in Figure 3 along with the scatter plots of the local response values and the corresponding local boxplots. The circles and the crosses inside the conditional central regions denote the conditional
Figure 3: The local response values (1st row), the local boxplots for the 1st (2nd row) and the 2nd coordinates (3rd row) and the conditional 50% central regions along with conditional medians (circle) and conditional 10% trimmed means (cross) (4th row) in the BudgetItaly Data.
Figure 4: The selected covariate curves (1st row), the local response values (2nd row), the local boxplots for the 1st (3rd row) and the 2nd coordinates (4th row) and the conditional 50% central regions along with conditional medians (circle) and conditional 10% trimmed means (cross) (5th row) in the Tecator Data.
medians and the 10% conditional trimmed means, respectively. Similar plots for the Tecator Data are presented in Figure 4.

The scatter plots of the local response values in Figure 3 and Figure 4 indicate the correlation between the two coordinates of the respective response variables. The correlation is higher in the case of the Tecator Data. The local boxplots and the conditional central regions both demonstrate the heteroscedasticity present in the two datasets. The local boxplots and the conditional medians and the conditional trimmed means reflect the variation of conditional skewness of the response over the covariate values in both the datasets.

3 Conditional Spread and Heteroscedasticity

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses. Some tests of heteroscedasticity were also proposed in the case of real valued response and covariate. Some tests of heteroscedasticity were also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.

We define a measure of conditional spread based on the conditional central region, and develop a test for heteroscedasticity in this section. In Liu et al. (1999), a measure of spread based on central regions of multivariate response as in our case. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity is developed in a nonparametric setup, where the response is multivariate and the covariate lies in a complete separable metric space. This covers the cases of real valued response as well as finite and infinite dimensional covariates. To the best of our knowledge, there is no other nonparametric test of heteroscedasticity for a multivariate response (Gupta and Tang, 1984; Holgersson and Shukur, 2004). Our test of heteroscedasticity was also proposed in the case of parametric regression with finite dimensional responses.
3.1 Data Demonstration

We now demonstrate the conditional spread measure in the BudgetItaly Data and the Tecator Data considered in Example 1 and Example 2, respectively, in section 1. For the BudgetItaly Data, we take the pair of food share and housing and fuels share as the response, and the variable pair of food price and housing and fuels price as the covariate. For the Tecator Data, we consider two regression problems. In the first case, we take the pair of fat and protein contents as a bivariate response, and in the second case, we take the triplet of moisture, fat, and protein contents as a trivariate response. The covariate is same in both cases, namely, the curve of absorbance spectra. For both the datasets, we compute $\Delta_n(r \mid x)$ with $r = 0.5$ and $x$ varying over the sample covariate values $X_i, i = 1,\ldots,n$. The choice of weights and the depth function are the same as in subsection 2.2. We compute the scores of the observations with respect to the first and the second principal components of the sample dispersion of the covariate, and denote the two scores as $P_1$ and $P_2$, respectively. For the BudgetItaly Data, we plot $\Delta_n(r \mid x)$ against $P_1$ and $P_2$ in Figure 5. For the Tecator Data, we plot $\Delta_n(r \mid x)$ for both the bivariate response and the trivariate response against $P_1$ and $P_2$ in Figure 6. We notice that the indication of heteroscedasticity in the BudgetItaly Data is not so strong in Figure 5, compared to that for the Tecator Data in Figure 6. In the Tecator Data, we can see clear pattern in the plots of $\Delta_n(r \mid x)$ against $P_1$ for both the bivariate and the trivariate responses in Figure 6, which indicates the presence of heteroscedasticity. The plots of $\Delta_n(r \mid x)$ against $P_2$ does not show much heteroscedasticity.

3.2 A Test for Heteroscedasticity

We now develop a nonparametric test for heteroscedasticity based on the conditional central region. In the presence of heteroscedasticity, $\Delta(r \mid x)$ will vary with $x$. To capture the variation of $\Delta(r \mid x)$ over $x$, we compute $\Delta_n(r \mid X_i)$ for $i = 1,\ldots,n$. Our hypotheses are

\[ H_0 : \Delta(r \mid x) \text{ is constant over } x, \]
\[ H_A : \Delta(r \mid x) \text{ varies with } x. \]
Figure 6: Plots of $\Delta_n(r \mid x)$ with $r = 0.5$ against $P_1$ and $P_2$ for bivariate and trivariate responses in the Tecator Data.
We define our test-statistic as

\[ T_n = \frac{1}{n} \sum_{i=1}^{n} \left[ \Delta_n(r \mid X_i) - \left( \frac{1}{n} \sum_{j=1}^{n} \Delta_n(r \mid X_j) \right) \right]^2. \]

Large values of \( T_n \) bear evidence against \( H_0 \). The p-value of the test is computed using a permutation procedure. Let \( S_n \) be the set of all permutations of the integers 1, ..., \( n \), defined as \( S_n = \{ \sigma : \{1, ..., n\} \to \{1, ..., n\}, \sigma \) is 1-1 and onto\}. Consider the set \( S_n = \{ (X_1, Y_{\sigma(1)}), ..., (X_n, Y_{\sigma(n)}) \mid \sigma \in S_n \} \), the set of all permuted samples, where the response is freely permuted. We compute the value of \( T_n \) for all the permuted samples in \( S_n \), and the empirical distribution of those values can be taken as an approximation of the null distribution of \( T_n \). The p-value is computed as the proportion of those values of \( T_n \) which are larger than the actually observed value of \( T_n \). In practice, the number of all permutations is too large for even moderate sample sizes, and we consider a fixed number of random permutations of 1, ..., \( n \) to calculate the p-value based on that. We present Table 1: p-values for the test of heteroscedasticity based on several conditional depths in the BudgetItaly Data and the Tecator Data

| Data      | Response type | Halfspace | Spatial | Projection | Simplicial |
|-----------|---------------|-----------|---------|------------|-----------|
| BudgetItaly | Bivariate     | 0.018     | 0.006   | 0          | 0.046     |
| Tecator   | Bivariate     | 0         | 0       | 0          | 0         |
| Tecator   | Trivariate    | 0.042     | 0       | 0.002      | 0         |

Table 1: p-values for the test of heteroscedasticity based on several conditional depths in the BudgetItaly Data and the Tecator Data

Overall, the p-values indicate strong presence of heteroscedasticity in both the data. We note that, though the plots of the conditional spread measure in the BudgetItaly Data in Figure 5 did not clearly indicate the presence of heteroscedasticity, our test detects the presence of heteroscedasticity in the data. The p-values for the Tecator Data are consistent with the plots in Figure 6 as both indicate presence of heteroscedasticity.

### 3.3 Level and Power study for the test of heteroscedasticity

We present a level and power study for our test of heteroscedasticity. We consider four regression models. Let \( \Sigma_p = ((\sigma_{ij}))_{p \times p} \) with \( \sigma_{ij} = 0.5 + 0.5I(i = j) \), and \( a \) be a positive number. In the first model, we consider a trivariate covariate and a bivariate response. The trivariate covariate \( X = (X^{(1)}, X^{(2)}, X^{(3)}) \) is such that \( X^{(1)}, X^{(2)} \) and \( X^{(3)} \) have independent \( Uniform[0, 1.5] \) distribution. The conditional distribution of the response \( Y \) given covariate \( X \) is bivariate normal with mean vector \( (0, 0) \) and dispersion matrix \((1 + a(X^{(1)}X^{(2)}X^{(3)}))\Sigma_2\). In the second model, the covariate is the same as in the first model, but the response \( Y \) is trivariate, with its conditional distribution given \( X \) being trivariate normal with mean vector \( (0, 0, 0) \) and dispersion matrix \((1 + a(X^{(1)}X^{(2)}X^{(3)}))\Sigma_3\). In the third and the fourth regression models, we consider a functional covariate \( X \), such that \( X(t) = Be^t \), with \( B \sim Uniform[0, 1] \) and \( t \in [0, 1] \). \( X \) is considered as a random element in \( L_2[0, 1] \). In the third model, we take the conditional distribution of the response \( Y \) given \( X \) to be bivariate normal with mean vector \( (0, 0) \) and dispersion matrix
Table 2: Estimated powers for bivariate and trivariate responses with trivariate covariate

| n  | level | Bivariate Y, trivariate X | Trivariate Y, trivariate X |
|----|-------|----------------------------|---------------------------|
| 100| 5%    | a = 0 0.052 0.322 0.576 0.724 0.800 | a = 0 0.054 0.098 0.184 0.254 0.348 |
| 100| 1%    | a = 2 0.008 0.142 0.318 0.486 0.574 | a = 2 0.006 0.022 0.070 0.114 0.170 |
| 200| 5%    | a = 4 0.046 0.718 0.948 0.986 0.996 | a = 4 0.052 0.330 0.588 0.760 0.858 |
| 200| 1%    | a = 6 0.008 0.500 0.844 0.936 0.976 | a = 6 0.010 0.146 0.390 0.536 0.660 |
| 300| 5%    | a = 8 0.044 0.896 0.998 1.000 1.000 | a = 8 0.070 0.626 0.936 0.992 0.998 |
| 300| 1%    | a = 0 0.014 0.734 0.976 1.000 1.000 | a = 0 0.012 0.374 0.778 0.924 0.980 |

Table 3: Estimated powers for bivariate and trivariate responses with functional covariate

| n  | level | Bivariate Y, functional X | Trivariate Y, functional X |
|----|-------|----------------------------|---------------------------|
| 100| 5%    | a = 0 0.056 0.706 0.740 0.752 0.760 | a = 0 0.050 0.316 0.376 0.400 0.400 |
| 100| 1%    | a = 2 0.006 0.436 0.458 0.476 0.488 | a = 2 0.004 0.110 0.150 0.172 0.172 |
| 200| 5%    | a = 4 0.038 0.982 0.996 0.994 0.994 | a = 4 0.052 0.750 0.786 0.796 0.800 |
| 200| 1%    | a = 6 0.008 0.896 0.916 0.928 0.928 | a = 6 0.004 0.512 0.560 0.588 0.586 |
| 300| 5%    | a = 8 0.052 0.998 0.998 1.000 1.000 | a = 8 0.048 0.934 0.958 0.966 0.972 |
| 300| 1%    | a = 0 0.008 0.984 0.992 0.996 0.994 | a = 0 0.008 0.722 0.772 0.786 0.786 |

$(1+a\|X\|_2)\Sigma_2$, where $\|\cdot\|_2$ is the $L_2$-norm. In the fourth model, the conditional distribution of the response $Y$ given $X$ is trivariate normal with mean vector $(0,0,0)$ and dispersion matrix $(1 + a\|X\|_2)\Sigma_3$. We estimate the power of the test of heteroscedasticity based on conditional halfspace depth for different sample sizes and different values of the constant $a$ based on 500 independent replications, with the number of random permutations in each replication being also 500. The estimated powers for the first and the second models are presented in Table 2, while those for the third and the fourth models are presented in Table 3. Note that the estimated power corresponding to $a = 0$ is nothing but the estimated size of the test. The observed power increases with the increment in the value of $a$, which is a parameter that controls the extent of heteroscedasticity in the model. Evidently, the power of the test tends to 1 as the sample size increases.

4 Asymptotic Results

Recall the definitions of $\mu(\cdot|z)$, $\mu_n(\cdot|x)$, $\rho(\cdot|x)$ and $\rho_n(\cdot|x)$ from section 2. We shall first establish the uniform asymptotic consistency for some standard conditional depth functions. We need the following conditions on the weights $W_{i,n}(x)$ and the conditional probability $\mu(\cdot|z)$ of $Y$ given $X = z$.

$$C(i,a) (\log n) \sum_{i=1}^n W_{i,n}^2(x) \overset{a.s.}{\rightarrow} 0$$

and

$$\max_{1 \leq i \leq n} W_{i,n}(x) \overset{a.s.}{\rightarrow} 0$$
as $n \rightarrow \infty$, and for any $\delta > 0$,

$$\sum_{i=1}^n W_{i,n}(x)I(d(x, X_i) \geq \delta) \overset{a.s.}{\rightarrow} 0$$
as $n \rightarrow \infty$. 

14
C(i.b) There is a collection of indices $S_n \subset \{1, 2, \ldots, n\}$ with cardinality $k_n$ such that for any $\delta > 0$, $S_n \subset \{i \mid d(x, X_i) < \delta, i = 1, \ldots, n\}$ for all sufficiently large $n$. Also,

$$\frac{\log n}{k_n} \to 0 \text{ as } n \to \infty,$$

$$\min_{i \in S_n} \frac{W_{i,n}(x)}{\max_{i \in S_n} W_{i,n}(x)} \text{ is almost surely bounded away from 0 as } n \to \infty,$$

and

$$\sum_{\{i_1, \ldots, i_{p+1}\} \not\subset S_n} W_{i_1,n}(x) \ldots W_{i_{p+1},n}(x) \to 0 \text{ as } n \to \infty.$$

C(ii) $\mu(\cdot \mid z) \overset{w}{\to} \mu(\cdot \mid \mathbf{x})$ as $z \to \mathbf{x}$.

Suppose that $P[d(x, \mathbf{X}) \leq h] > 0$ for all $h > 0$. Then, in the nearest neighbor method of constructing weights $W_{i,n}(x)$, conditions C(i.a) and C(i.b) hold when $k_n = \lfloor \log n \rfloor + 1$. In the kernel method, suppose that the kernel function $K(\cdot)$ satisfies $c(0 \leq s \leq 1) \leq K(s) \leq uI(0 \leq s \leq 1)$ for some constants $0 < u < \infty$. Depending on the distribution of the covariate $X$, we can choose a sequence of bandwidths $\{h_n\}$ such that $h_n \to 0$ and $[nP[d(x, \mathbf{X}) \leq h_n]]^{-1}(\log n) \to 0$ as $n \to \infty$. For this choice of $\{h_n\}$ and the kernel function, it can be shown using the Bernstein inequality (see Serfling (2009, p. 95, Lemma A)) and the Borel-Cantelli Lemma that conditions C(i.a) and C(i.b) are satisfied.

Assumption C(ii) holds in many common situations that we encounter. As an example, consider the location-scale model defined by $Y = l(\mathbf{X}) + s(\mathbf{X})G$. Here, $\mathbf{G}$ is a random vector independent of $\mathbf{X}$, and the functions $l(\cdot) : \mathcal{C} \to \mathbb{R}^p$ and $s(\cdot) : \mathcal{C} \to \mathbb{R}$ are both continuous at $x$. In such a setup, $\mu(\cdot \mid \mathbf{z}) \overset{w}{\to} \mu(\cdot \mid \mathbf{x})$ as $\mathbf{z} \to \mathbf{x}$. Under C(i.a) and C(ii) it follows that

$$\mu_n(\cdot \mid \mathbf{x}) \overset{w}{\to} \mu(\cdot \mid \mathbf{x}) \text{ almost surely as } n \to \infty. \tag{4.1}$$

See Lemma B.1 for the proof of (4.1). The following theorem states the uniform consistency for several depth functions.

**Theorem 4.1.** Let $\rho(y \mid \mathbf{x})$ be any of the four conditional depth functions described in section 2. Suppose that there is a neighborhood of $\mathbf{x}$ such that for all $\mathbf{z}$ in that neighborhood, the conditional distribution of $Y$ given $\mathbf{X} = \mathbf{z}$ has a continuous positive density $f(\cdot \mid \mathbf{z})$ that is continuous in $\mathbf{z}$. Then, under conditions C(i.a), C(i.b) and C(ii),

$$\sup_{y \in \mathbb{R}^p} |\rho_n(y \mid \mathbf{x}) - \rho(y \mid \mathbf{x})| \overset{P}{\to} 0 \text{ as } n \to \infty. \tag{4.2}$$

Examples of continuous positive conditional densities $f(\cdot \mid \mathbf{z})$ that are continuous in $\mathbf{z}$ include the location-scale model: $Y = l(\mathbf{X}) + s(\mathbf{X})G$, where $G$ is a random vector independent of $\mathbf{X}$ and has a continuous positive density on $\mathbb{R}^p$, and the functions $l(\cdot) : \mathcal{C} \to \mathbb{R}^p$ and $s(\cdot) : \mathcal{C} \to \mathbb{R}$ are both continuous in a neighborhood of $\mathbf{x}$.

### 4.1 Asymptotic consistency of conditional central regions and related measures

We now proceed to state the asymptotic consistency of the conditional central regions, medians and trimmed means, and spread measures. Recall the definitions of $D(\cdot \mid \mathbf{x})$, $D_n(\cdot \mid \mathbf{x})$, $\alpha(r)$, $\alpha_n(r)$, $\mathbf{m}(\mathbf{x})$, $\mathbf{m}_n(\mathbf{x})$, $\mathbf{m}(r \mid \mathbf{x})$ and $\mathbf{m}_n(r \mid \mathbf{x})$ from section 2. We need the following conditions.
C(iii) \( \mu(\{y \mid \rho(y \mid x) = \alpha \} \mid x) = 0 \) for all \( \alpha \).

C(iv) For any \( 0 \leq \gamma_1 < \gamma_2 \leq \rho(m(x) \mid x) \), \( \mu(\{y \mid \gamma_1 \leq \rho(y \mid x) \leq \gamma_2 \} \mid x) > 0 \).

C(v) \( \rho(y \mid x) \) is a continuous function of \( y \), and \( \rho(y \mid x) \to 0 \) as \( \|y\| \to \infty \).

C(vi) Define \( D_0(\alpha \mid x) = \{y \mid \rho(y \mid x) > \alpha \} \). Then, the closure of \( D_0(\alpha \mid x) \) is \( D(\alpha \mid x) \) for any \( 0 < \alpha < \rho(m(x) \mid x) \).

Assumptions C(iii) and C(iv) imply that the distribution function of the random variable \( \rho(Y \mid x) \) is continuous and strictly increasing. Assumptions C(v) and C(vi) are related to the smoothness of the conditional depth and the corresponding central regions.

Suppose \( \mu(\cdot \mid x) \) has a probability density with a convex support. Then, from the proof of Lemma 6.1 in [Donoho and Gasko (1992)], we get that C(vi) is satisfied for the conditional halfspace depth. From the arguments in the proof of Lemma 6.3 in Donoho and Gasko (1992), we get that \( D(\alpha \mid x) \) is convex for all \( 0 < \alpha < \rho(m(x) \mid x) \), and consequently, C(iii), C(iv) and C(vi) are also satisfied for the conditional halfspace depth.

Now suppose that \( \mu(\cdot \mid x) \) has a positive probability density on \( \mathbb{R}^p \). Then, the assumptions C(iii), C(iv) and C(v) hold for the conditional spatial depth. Note that when assumption C(v) holds, assumption C(vi) is also satisfied if the conditional depth has no local maximum. From the arguments in section 2 in Chowdhury and Chaudhuri (2016), we get that the conditional spatial median is unique. Using arguments similar to those in the proofs of Lemma 2.5 and Lemma 2.6 in the supplement of Chowdhury and Chaudhuri (2016), one can show that the conditional spatial depth \( \rho(y \mid x) \) has non-zero Fréchet derivative with respect to \( y \) for \( y \) not being the conditional spatial median. Consequently, the conditional spatial depth cannot have any local maximum and hence it satisfies assumption C(vi).

From Theorem 3.4 in Zuo and Serfling (2000b), it follows that it is sufficient for \( \mu(\cdot \mid x) \) to have an elliptically symmetric probability density on \( \mathbb{R}^p \) for the associated conditional projection depth and conditional simplicial depth to satisfy assumptions C(iii), C(iv), C(v) and C(vi).

In the theorem below, we establish an asymptotic consistency result for the conditional sample central regions. Note that the Hausdorff distance \( d_H(A, B) \) between two nonempty closed and bounded subsets of a metric space is defined as \( d_H(A, B) = \inf \{\epsilon \mid A \subseteq B^\epsilon, B \subseteq A^\epsilon\} \), where \( A^\epsilon \) and \( B^\epsilon \) denote the \( \epsilon \)-neighborhoods of \( A \) and \( B \), respectively (see, e.g., Munkres (2000, p. 281)).

**Theorem 4.2.** Suppose C(iii), C(iv), (4.1) and (4.2) are satisfied. Then, for any \( \epsilon > 0 \) and any \( 0 < r < 1 \),

\[
P[D(\alpha(r) + \epsilon \mid x) \subseteq D_n(\alpha_n(r) \mid x) \subseteq D(\alpha(r) - \epsilon \mid x)] \to 1 \quad \text{as} \ n \to \infty.
\]

If in addition, conditions C(v) and C(vi) hold, then \( d_H(D_n(\alpha_n(r) \mid x), D(\alpha(r) \mid x)) \to 0 \) as \( n \to \infty \) for any \( 0 < r < 1 \).

The conditional \( \alpha \)-depth contour \( \delta(\alpha \mid x) \) of \( Y \) given \( X = x \) is defined as the boundary of \( D(\alpha \mid x) \), and for \( 0 < r < 1 \), the conditional \( 100r\% \) central region contour of \( Y \) given \( X = x \) is defined as \( \delta(\alpha(r) \mid x) \). The sample counterparts of \( \delta(\alpha \mid x) \) and \( \delta(\alpha(r) \mid x) \) are denoted as \( \delta_n(\alpha \mid x) \) and \( \delta_n(\alpha_n(r) \mid x) \), respectively. A consequence of Theorem 4.2 stated below in Corollary 4.2.1 is that for any band around \( \delta(\alpha(r) \mid x) \), which may be of arbitrarily small width, \( \delta_n(\alpha_n(r) \mid x) \) eventually lies inside that band with high probability.
for large $n$. Since the contours determine the shapes of the central regions, this implies that the shapes of the sample central regions are good approximations of their population counter-parts in large samples.

**Corollary 4.2.1.** Suppose $C(iii)$, $C(iv)$, (4.1) and (4.2) are satisfied. Then, for any $\epsilon > 0$,

$$P \left[ \delta_n(\alpha_n(r) \mid x) \subseteq \{ y \in \mathbb{R}^p \mid \alpha(r) - \epsilon \leq \rho(y \mid x) < \alpha(r) + \epsilon \} \right] \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$ 

Convergence of sample central regions were studied earlier in the unconditional setup by He and Wang (1997) and Zuo and Serfling (2000b). In He and Wang (1997), the authors assumed that the population depth contours are elliptic in nature in order to establish the convergence of the contours of the central regions. In Zuo and Serfling (2000b), the authors restricted themselves to elliptic distributions for proving the convergence of the central regions in the Hausdorff distance.

The next theorem states the asymptotic consistency of the sample conditional median and the sample conditional trimmed means.

**Theorem 4.3.** Let $M(x)$ be the set of all conditional medians corresponding to the population conditional depth function $\rho(y \mid x)$. Suppose that conditions $C(iii)$, $C(vi)$ are satisfied and (4.1) and (4.2) hold. Then, for any sequence of sample conditional medians $\{m_n(x)\}$,

$$\inf_{m \in M(x)} \| m_n(x) - m \| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$ 

Also, $m_n(r \mid x) \rightarrow m(r \mid x)$ as $n \rightarrow \infty$ for any $0 < r < 1$.

We state the asymptotic consistency of the conditional spread estimates $S_n(r \mid x)$ and $\Delta_n(r \mid x)$ in the theorem below.

**Theorem 4.4.** (a) Suppose $C(iii)$, $C(iv)$, (4.1) and (4.2) hold, and $\mu(\cdot \mid x)$ has a continuous positive density on $\mathbb{R}^p$. Then, $S_n(r \mid x) \rightarrow S(r \mid x)$ as $n \rightarrow \infty$.

(b) Let $0 < r < 1$, and suppose that for any open set $G$ with $G \cap D(\alpha(r) \mid x) \neq \emptyset$, we have $P \left[ Y \in G \cap D(\alpha(r) \mid x) \right] > 0$. Then, under $C(iii)$, $C(vi)$, (4.1) and (4.2), $\Delta_n(r \mid x) \rightarrow \Delta(r \mid x)$ as $n \rightarrow \infty$.

The first condition in part (b) in the above theorem is trivially satisfied for any depth function when the support of the distribution of $Y$ is the whole response space. On the other hand, for particular depths like the conditional halfspace depth, the conditional central regions are bounded convex sets. So, if the support of the distribution of $Y$ is a bounded convex set, then also that condition is satisfied for the conditional halfspace depth.

### A Proofs and Mathematical Details

Here, we provide the proofs of the theorems in section 4. The proofs here utilize some lemmas, which are presented in Appendix B.

**Proof of Theorem 4.4.** Let $\rho(y \mid x)$ be the conditional halfspace depth. Define $H(u, y) = \{ v \in \mathbb{R}^p \mid u'v \leq u' y \}$. From (6.6) in Donoho and Gasko (1992), we get $\sup_{y \in \mathbb{R}^p} | \rho_n(y \mid x) - \rho(y \mid x) | \leq \sup \{ | \mu_n(H(u, y) \mid x) - \mu(H(u, y) \mid x) | \mid u, y \in \mathbb{R}^p \}$. Also, the class of sets $\{ H(u, y) \mid u, y \in \mathbb{R}^p \}$ forms a VC-class (Van der Vaart and Wellner 2000 p. 152). Under the conditions in the theorem, every $H(u, y)$ is a $\mu(\cdot \mid x)$-continuity set, and from Lemma
we get that given any \( \epsilon > 0, P[\sup_{\|y\| \leq C}|\mu_n(H(u, y) | x) - \mu(H(u, y) | x)| | u, y \in \mathbb{R}^p] > \epsilon |X_1, X_2, \ldots| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Consequently, from an application of the dominated convergence theorem, we get that (4.2) is satisfied.

Next, let \( \rho(y | x) \) be the conditional spatial depth. Define \( Q(y | z) = E[\|y - Y\|^{-1}(y - Y) | X = z] \). Under the conditions in the theorem, we have \( Q(y | z) \) continuous at \( z = x \). Further, \( \sup_{y \in \mathbb{R}^p} |\rho_n(y | x) - \rho(y | x)| \leq \sup_{y \in \mathbb{R}^p} |Q_n(y | x) - Q(y | x)| \), where \( Q_n(y | x) = \sum_{i=1}^n \|y - Y\|^{-1}(y - Y) W_{i,n}(x) \). Under conditions C(i.a) and C(ii) the continuity of \( Q(y | z) \) at \( z = x \), using arguments similar to those in the proof of Lemma B.3, and some arguments related to the properties of VC-subgraph classes similar to those in the proof of Theorem 5.5 in Koltchinskii (1997), we get that given any \( \epsilon > 0, P[\sup_{y \in \mathbb{R}^p} |Q_n(y | x) - Q(y | x)| > \epsilon |X_1, X_2, \ldots| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Again, from an application of the dominated convergence theorem, we get that (4.2) is satisfied.

Now, let \( \rho_n(y | x) \) denote the estimate of \( \rho(y | x) \) constructed using only the sample observations corresponding to the indices in \( S_n \). Then, under condition C(i.b), \( \sup_{y \in \mathbb{R}^p} |\rho_n(y | x) - \rho_n(y | x)| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Consequently, for any \( \epsilon > 0 \), we have \( P[|\rho_n(y | x) - \rho(y | x)| > \epsilon |X_1, \ldots, X_n| \leq P[|\rho_n(y | x) - \rho(y | x)| > \epsilon/2 |X_1, \ldots, X_n| \) for all sufficiently large \( n \) almost surely. Under the conditions in the theorem, every \( p \)-dimensional simplex is a \( \mu(\cdot | x) \)-continuity set. Consequently, for any \( y \in \mathbb{R}^p \) and given any \( \epsilon > 0 \), under conditions C(i.b) and C(ii) using arguments similar to those used in subsection 5a in Hoeffding (1963) and in the proof of Lemma B.1, we can get an exponential upper bound for \( P[|\rho_n(y | x) - \rho(y | x)| > \epsilon/2 |X_1, \ldots, X_n| \). Using that upper bound, the fact that the class of sets \( \{C_y \ y \in \mathbb{R}^p \} \), where \( C_y = \{(y_1, \ldots, y_{p+1}) | y \in \Delta(y_1, \ldots, y_{p+1}) \} \), is a VC-class of sets (Arcones and Gine, 1993, Corollary 6.7, p. 1539), and arguments similar to those used for Corollary 3.3 in Arcones and Gine (1993, p. 1512), it follows that \( P[\sup_{y \in \mathbb{R}^p} |\rho_n(y | x) - \rho(y | x)| > \epsilon |X_1, X_2, \ldots| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Again, from an application of the dominated convergence theorem, we get that (4.2) is satisfied.

Finally, we consider \( \rho(y | x) \) to be the conditional projection depth. Under the conditions in the theorem, we get that for all sufficiently large \( n \), \( \sup_{|y| > C} \rho_n(y | x) \xrightarrow{a.s.} 0 \) as \( C \to \infty \), and \( \sup_{|y| > C} \rho(y | x) \to 0 \) as \( C \to \infty \) (see Lemma B.2 for proof). Hence, it is sufficient to show that for any \( C > 0 \), \( \sup_{|y| \leq C} |\rho_n(y | x) - \rho(y | x)| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Under the conditions in the theorem, we also get that

\[
\sup_{|y| \leq C} \sup_{|u| = 1} \frac{|u^t y - M_n(u^t Y | x)|}{M_n(|u^t Y - M(u^t Y | x)| | x)} - \sup_{|u| = 1} \frac{|u^t y - M(u^t Y | x)|}{M(|u^t Y - M(u^t Y | x)| | x)} \xrightarrow{a.s.} 0
\]

as \( n \to \infty \) (see Lemma B.2 for proof), and this implies \( \sup_{|y| \leq C} |\rho_n(y | x) - \rho(y | x)| \xrightarrow{a.s.} 0 \) as \( n \to \infty \). Hence, (4.2) holds for the conditional projection depth.

**Proof of Theorem 4.2** When (4.2) holds, one can show that for any \( \epsilon, \delta > 0 \) and any sequence \( \{\alpha_n\} \) with \( \alpha_n \xrightarrow{P} \alpha \) as \( n \to \infty \), we have

\[
P[D(\alpha + \delta | x) \subseteq D_n(\alpha_n | x) \subseteq D(\alpha - \delta | x)] > 1 - \epsilon
\]

for all sufficiently large \( n \) (see Lemma B.3 for proof). Also, one can show that under C(iii) C(iv) (4.1), (4.2) and for any \( 0 < r < 1 \),

\[
\alpha_n(r) \xrightarrow{P} \alpha(r) \text{ as } n \to \infty.
\]
See Lemma B.5 for a proof of (A.2). From (A.1) and (A.2), we get that for any \( \epsilon > 0 \) and any \( 0 < r < 1 \),

\[
P[D(\alpha(r) + \epsilon \mid x) \subseteq D_n(\alpha_n(r) \mid x) \subseteq D(\alpha(r) - \epsilon \mid x)] \to 1 \tag{A.3}
\]

as \( n \to \infty \). Next, under \( [C(v)] \) \( [C(vi)] \) and (4.2), and for any sequence of random variables \( \{\alpha_n\} \) with \( \alpha_n \overset{p}{\to} \alpha \) as \( n \to \infty \), we get \( d_H(D_n(\alpha_n \mid x), D(\alpha \mid x)) \overset{p}{\to} 0 \) as \( n \to \infty \) (see Lemma B.6 for proof). From this and (A.2), we have \( d_H(D_n(\alpha_n(r) \mid x), D(\alpha(r) \mid x)) \overset{p}{\to} 0 \) as \( n \to \infty \).

\[\Box\]

**Proof of Corollary 4.2.1** The proof follows directly from (A.3) in the proof of Theorem 4.2.

**Proof of Theorem 4.3.** We denote \( \rho(M(x) \mid x) = \max_y \rho(y \mid x) = \rho(m \mid x) \) for any \( m \in M(x) \). From (4.2), we get that for any \( \epsilon, \delta > 0 \),

\[
P[\rho_n(m_n(x) \mid x) \leq \rho(m_n(x) \mid x) + \epsilon \leq \rho(M(x) \mid x) + \epsilon] > 1 - \frac{\delta}{2},
\]

\[
P[\rho_n(m_n(x) \mid x) \geq \rho_n(M(x) \mid x) \geq \rho(M(x) \mid x) - \epsilon] > 1 - \frac{\delta}{2}
\]

for all sufficiently large \( n \), which implies that \( \rho_n(m_n(x) \mid x) \overset{p}{\to} \rho(M(x) \mid x) \) as \( n \to \infty \). From this fact and (A.1), we have, given any \( \epsilon, \delta > 0 \),

\[
P[D_n(\rho_n(m_n(x) \mid x) \mid x) \subseteq D(\rho(M(x) \mid x) - \delta \mid x)] > 1 - \epsilon
\]

for all sufficiently large \( n \). Note that \( D_n(\rho_n(m_n(x) \mid x) \mid x) \) is the collection of all sample conditional medians, and \( M(x) = D(\rho(M(x) \mid x) \mid x) \subseteq D(\rho(M(x) \mid x) - \delta \mid x) \) for all \( \delta > 0 \). Whenever \( D_n(\rho_n(m_n(x) \mid x) \mid x) \subseteq D(\rho(M(x) \mid x) - \delta \mid x) \), we have

\[
d_H(D_n(\rho_n(m_n(x) \mid x) \mid x), M(x)) \leq d_H(D(\rho(M(x) \mid x) - \delta \mid x), M(x)).
\]

From (B.18) in the proof of Lemma B.6 we get

\[
\lim_{\delta \to 0^+} d_H(M(x), D(\rho(M(x) \mid x) - \delta \mid x)) = 0.
\]

Therefore, \( d_H(D_n(\rho_n(m_n(x) \mid x) \mid x), M(x)) \overset{p}{\to} 0 \) as \( n \to \infty \), which implies \( \inf_{m \in M(x)} \|m_n(x) - m\| \overset{p}{\to} 0 \) as \( n \to \infty \) for any sequence of sample conditional medians \( \{m_n\} \).

We now proceed to prove the consistency of \( m_n(r \mid x) \). Let \( \{\alpha_n\} \) be a sequence of random variables with \( \alpha_n \overset{p}{\to} \alpha > 0 \) as \( n \to \infty \). Suppose that \( [C(iii)] \) \( [C(v)] \) (4.1) and (4.2) are satisfied. Then,

\[
\int y I(y \in D_n(\alpha_n \mid x)) \mu_n(dy \mid x) \overset{p}{\to} \int y I(y \in D(\alpha \mid x)) \mu(dy \mid x) \tag{A.4}
\]

as \( n \to \infty \) (see Lemma B.7 for proof). For any \( 0 < r < 1 \), from conditions \( [C(iii)] \) and \( [C(iv)] \) we get that \( \alpha(1-r) > 0 \). So, from (A.2) and (A.4), we have \( \int y I(y \in D_n(\alpha_n(1-r) \mid x)) \mu_n(dy \mid x) \overset{p}{\to} \int y I(y \in D(\alpha(1-r) \mid x)) \mu(dy \mid x) \) as \( n \to \infty \). Also, from \( [C(iii)] \) (4.1) and Theorem 4.2 it follows that \( \mu_n(D_n(\alpha_n(1-r) \mid x) \mid x) \overset{p}{\to} \mu(D(\alpha(1-r) \mid x) \mid x) \) as \( n \to \infty \). Hence, \( m_n(r \mid x) \overset{p}{\to} m(r \mid x) \) as \( n \to \infty \). \( \Box \)
Proof of Theorem 4.4. Proof of (a): Suppose \( C(iii) \) and \((4.2)\) are satisfied, and \( \mu(\cdot | x) \) has a continuous positive density on \( \mathbb{R}^p \). Then, for any sequence \( \{\alpha_n\} \) with \( \alpha_n \xrightarrow{p} \alpha > 0 \) as \( n \to \infty \), we have\( \lambda(D_n(\alpha_n | x)) \xrightarrow{p} \lambda(D(\alpha | x)) \) as \( n \to \infty \), where \( \lambda(\cdot) \) is the Lebesgue measure on \( \mathbb{R}^p \) (see Lemma B.8 for proof). Using this fact and \((A.2)\), the proof of part (a) is complete.

Proof of (b): The diameter of a set \( A \subset \mathbb{R}^p \) is defined as \( \text{Diameter}(A) = \sup\{\|u - v\| | u, v \in A\} \). For any pair of sets \( A \) and \( B \), \( |\text{Diameter}(A) - \text{Diameter}(B)| \leq 2d_H(A, B) \) (see Lemma B.9 for proof). Define \( D_n'(\alpha_n | x) = \{Y_i | \rho_n(Y_i | x) \geq \alpha_n; i = 1, \ldots, n\} \). Note that \( \Delta_n(r | x) = \text{Diameter}(D_n'(\alpha_n(r) | x)) \) and \( \Delta(r | x) = \text{Diameter}(D(\alpha(r) | x)) \). Let \( \{\alpha_n\} \) be a sequence of random variables with \( \alpha_n \xrightarrow{p} \alpha \) as \( n \to \infty \). Suppose that for any open set \( G \) with \( G \cap D(\alpha | x) \neq \emptyset \), we have \( P[ Y \in G \cap D(\alpha | x) ] > 0 \). Then, under \( C(v), C(vi) \) and \((4.2)\), \( d_H(D_n'(\alpha_n | x), D(\alpha | x)) \xrightarrow{p} 0 \) as \( n \to \infty \) (see Lemma B.10 for proof). From this fact and \((A.2)\), the proof of part (b) is complete.

B Additional Mathematical Details

Lemma B.1. Under \( C(i.a) \) and \( C(ii) \) for any \( \mu(\cdot | x) \)-continuity set \( B \), and given any \( k > 0 \) and \( \epsilon > 0 \), we have

\[
P[|\mu_n(B | x) - \mu(B | x)| > \epsilon | X_1, X_2, ...] < 2n^{-k}
\]

almost surely for all sufficiently large \( n \). Further, if \( B \) is a VC class of \( \mu(\cdot | x) \)-continuity sets, then

\[
P[\sup\{|\mu_n(B | x) - \mu(B | x)| | B \in B\} > \epsilon | X_1, X_2, ...] \xrightarrow{a.s.} 0 \quad \text{as} \quad n \to \infty.
\]

Proof. For any \( \epsilon > 0 \),

\[
P[|\mu_n(B | x) - \mu(B | x)| > \epsilon | X_1, X_2, ...]
\]

\[
\leq P\left[\sum_{i=1}^{n}\left[I(Y_i \in B) - \mu(B | X_i)\right] W_{i,n}(x) > \frac{\epsilon}{2} | X_1, X_2, ...\right] + P\left[\sum_{i=1}^{n}\left[\mu(B | X_i) - \mu(B | x)\right] W_{i,n}(x) > \frac{\epsilon}{2} | X_1, X_2, ...\right].
\]

Now, from the Bernstein inequality (see Serfling (2009) p. 95, Lemma A)), we have

\[
P\left[\sum_{i=1}^{n}\left[I(Y_i \in B) - \mu(B | X_i)\right] W_{i,n}(x) > \frac{\epsilon}{2} | X_1, X_2, ...\right] \leq 2 \exp\left[-\frac{\epsilon^2}{2 \sum_{i=1}^{n} W_{i,n}^2(x) + 2\epsilon \max_{1 \leq i \leq n} W_{i,n}(x)}\right].
\]

From \( C(i.a) \) we get that there is an integer \( N_1 \) such that for all \( n \geq N_1 \),

\[
(\log n) \sum_{i=1}^{n} W_{i,n}^2(x) < \frac{\epsilon^2}{4k} \quad \text{and} \quad (\log n) \max_{1 \leq i \leq n} W_{i,n}(x) < \frac{\epsilon}{4k}
\]

20
almost surely. \( N_1 \) depends on \( X_1, X_2, \ldots, k \) and \( \epsilon \). So, for all \( n \geq N_1 \),

\[
P \left[ \left| \sum_{i=1}^{n} \left[ I(Y_i \in B) - \mu(B \mid X_i) \right] W_{i,n}(x) \right| > \frac{\epsilon}{2} \left| X_1, X_2, \ldots \right| \right] < 2 \exp \left[ -k \log n \right]
\]

almost surely. On the other hand, for any \( \delta > 0 \), we have

\[
P \left[ \left| \sum_{i=1}^{n} \left[ \mu(B \mid X_i) - \mu(B \mid x) \right] W_{i,n}(x) \right| > \frac{\epsilon}{2} \left| X_1, X_2, \ldots \right| \right] \leq P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| I(d(x, X_i) \geq \delta) W_{i,n}(x) \right| > \frac{\epsilon}{4} \left| X_1, X_2, \ldots \right| \right] + P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| I(d(x, X_i) < \delta) W_{i,n}(x) \right| > \frac{\epsilon}{4} \left| X_1, X_2, \ldots \right| \right].
\]

Since \( B \) is a \( \mu(\cdot \mid x) \)-continuity set, from \( C(ii) \) we can find \( \delta > 0 \) such that \( |\mu(B \mid z) - \mu(B \mid x)| < (\epsilon/5) \) whenever \( d(x, z) < \delta \). From \( C(i.a) \) we can find an integer \( N_2 \) such that for all \( n \geq N_2 \),

\[
\sum_{i=1}^{n} W_{i,n}(x) I(d(x, X_i) \geq \delta) < \frac{\epsilon}{5}
\]

almost surely. \( N_2 \) depends on \( X_1, X_2, \ldots, \delta \) and \( \epsilon \). Therefore, for all \( n \geq N_2 \),

\[
P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| I(d(x, X_i) \geq \delta) W_{i,n}(x) \right| > \frac{\epsilon}{4} \left| X_1, X_2, \ldots \right| \right] = 0 \quad \text{almost surely},
\]

and

\[
P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| I(d(x, X_i) < \delta) W_{i,n}(x) \right| > \frac{\epsilon}{4} \left| X_1, X_2, \ldots \right| \right] \leq P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| I(d(x, X_i) < \delta) W_{i,n}(x) \right| > \frac{\epsilon}{4} \left| X_1, X_2, \ldots \right| \right] = 0
\]

almost surely. Hence, for all \( n \geq N_2 \),

\[
P \left[ \sum_{i=1}^{n} \left| \mu(B \mid X_i) - \mu(B \mid x) \right| W_{i,n}(x) \right| > \frac{\epsilon}{2} \left| X_1, X_2, \ldots \right| \right] = 0
\]

almost surely. Therefore, for all \( n \geq \max\{N_1, N_2\} \),

\[
P \left[ |\mu_n(B \mid x) - \mu(B \mid x)| > \epsilon \left| X_1, X_2, \ldots \right| \right] < 2^{-k}
\]  \hspace{1cm} (B.1)

almost surely for all sufficiently large \( n \).

Now, if \( B \) is a VC class of \( \mu(\cdot \mid x) \)-continuity sets, using (B.1) and some standard arguments involving the properties of VC-classes (see, e.g., Pollard (1984, p. 13–24)), we get

\[
P \left[ \sup \{ |\mu_n(B \mid x) - \mu(B \mid x)| \mid B \in B \} > \epsilon \left| X_1, X_2, \ldots \right| \right] \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty.
\]

\[\square\]
The following lemma is used in the proof of Theorem 4.1.

Lemma B.2. Let \( \rho(y \mid x) \) be the conditional projection depth function described in Section 2. Suppose that there is a neighborhood of \( x \) such that for all \( z \) in that neighborhood, the conditional distribution of \( Y \) given \( X = z \) has a continuous positive density \( f(\cdot \mid z) \) on \( \mathbb{R}^p \), which is continuous in \( z \). Then, under \( C(i.a), C(i.b) \) and \( C(ii) \), we have for all sufficiently large \( n \), \( \sup_{\|y\| > C} \rho_n(y \mid x) \overset{a.s.}{\to} 0 \) as \( C \to \infty \), and \( \sup_{\|y\| > C} \rho(y \mid x) \to 0 \) as \( C \to \infty \). Further,

\[
\sup_{\|y\| \leq C} \left( \sup_{\|u\| = 1} \frac{M_n(u^t y - M_n(u^t Y \mid x))}{M_n(u^t Y - M_n(u^t Y \mid x))} \right) \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to \infty.
\]

Here, \( M(u^t Y \mid x) \) and \( M((u^t Y - M(u^t Y \mid x)) \mid x) \) are the conditional medians of \( u^t Y \) and \( M((u^t Y - M(u^t Y \mid x)) \mid x) \) given \( X = x \), respectively. And \( M_n(u^t Y \mid x) \) and \( M_n((u^t Y - M(u^t Y \mid x)) \mid x) \) are the sample analogues of \( M(u^t Y \mid x) \) and \( M((u^t Y - M(u^t Y \mid x)) \mid x) \), respectively.

Proof. Under the condition in the theorem on the conditional density of \( Y \) given \( X = z \), we get that there is a neighborhood of \( x \) such that for \( z \) lying in that neighborhood and for every \( u \), \( u^t Y \) given \( X = z \) has a continuous strictly increasing conditional distribution function, which is also continuous as a function of \( u \). Consequently, \( M(u^t Y \mid x) \) and \( M((u^t Y - M(u^t Y \mid x)) \mid x) \) are unique for every \( u \) and continuous as functions of \( u \). Denote the conditional distribution function of \( u^t Y \) given \( X = x \) as \( F(\cdot \mid u, x) \). Let \( F_n(\cdot \mid u, x) \) be the corresponding weighted empirical distribution function of \( u^t Y \), which is defined by

\[
F_n(v \mid u, x) = \sum_{i=1}^{n} I(u^t Y_i \leq v) W_{i,n}(x).
\]

Under the condition in the theorem on the conditional density of \( Y \) given \( X = z \), using \( C(i.a) \) and \( C(ii) \) and arguments similar to those in the proof of Lemma B.1, we get that \( F_n(v \mid u, x) \overset{a.s.}{\to} F(v \mid u, x) \) as \( n \to \infty \) for all \( v \in \mathbb{R} \) and \( u \in \mathbb{R}^p \). Recall that a halfspace in \( \mathbb{R}^p \) is defined as \( H(u, v) = \{ y \in \mathbb{R}^p \mid u^t y \leq v \} \), where \( u \in \mathbb{R}^p \) and \( v \in \mathbb{R} \). Since the class of all halfspaces \( \{ H(u, v) \mid u \in \mathbb{R}^p, v \in \mathbb{R} \} \) forms a VC class (see Van der Vaart and Wellner (2000, p. 152)), using some standard arguments involving the properties of VC-classes (see, e.g., Pollard (1984, p. 13–24)), we get that

\[
\sup_{u \in \mathbb{R}^p, v \in \mathbb{R}} |F_n(v \mid u, x) - F(v \mid u, x)| \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to \infty. \tag{B.2}
\]

Now, from (B.2) and \( C(i.a) \), we have

\[
\sup_{\|u\| = 1} \left| F(M_n(u^t Y \mid x) \mid u, x) - F(M(u^t Y \mid x) \mid u, x) \right| \\
\leq \sup_{\|u\| = 1} \left| F_n(M_n(u^t Y \mid x) \mid u, x) - 0.5 \right| + \sup_{\|u\| = 1} \left| F_n(M_n(u^t Y \mid x) \mid u, x) - F(M_n(u^t Y \mid x) \mid u, x) \right| \\
\leq \max_{1 \leq i \leq n} W_{i,n}(x) + \sup_{u \in \mathbb{R}^p, v \in \mathbb{R}} |F_n(v \mid u, x) - F(v \mid u, x)| \overset{a.s.}{\to} 0 \quad \text{as} \quad n \to \infty. \tag{B.3}
\]

Since \( F(\cdot \mid u, x) \) is a continuous and strictly increasing distribution function, its inverse \( F^{-1}(\cdot \mid u, x) \) is well-defined. Further, under the condition in the theorem on the conditional
density of $Y$ given $X = z$, $F^{-1}(\cdot | u, x)$ is also continuous in $u$. Since $\{u \in \mathbb{R}^p | \|u\| = 1\}$ is compact, $F^{-1}(\cdot | u, x)$ is uniformly continuous in $u \in \{u \in \mathbb{R}^p | \|u\| = 1\}$. Hence, from (B.3), we get

$$
\sup_{\|u\|=1} |M_n(u'(Y | x) - M(u'(Y | x))|
= \sup_{\|u\|=1} |F^{-1}(F(M_n(u'(Y | x) | u, x) | x, x) - F^{-1}(F(M(u'(Y | x) | u, x) | u, x) |
\xrightarrow{a.s.} 0 \text{ as } n \to \infty.
$$

One can similarly show that

$$
\sup_{\|u\|=1} |M_n(u'(Y - M_n(u'(Y | x)) | x) - M(u'(Y - M(u'(Y | x)) | x)| \xrightarrow{a.s.} 0
$$
as $n \to \infty$. Now, for all sufficiently large $n$,

$$
\sup_{\|u\|=1} \frac{|u'y - M_n(u'y | x)|}{M_n(|u'y - M_n(u'y | x)| | x)}
\geq \sup_{\|u\|=1} \frac{|u'y - M_n(u'y | x)|}{M_n(|u'y - M_n(u'y | x)| | x)}
\sup_{\|u\|=1} \frac{\|y\| - \sup_{\|u\|=1} M_n(u'y | x)}{M_n(|u'y - M_n(u'y | x)| | x)} \xrightarrow{a.s.} \infty \text{ as } \|y\| \to \infty.
$$

So, for all sufficiently large $n$, $\sup_{\|y\| \leq C} \rho_n(y | x) \xrightarrow{a.s.} 0$ as $C \to \infty$, and $\sup_{\|y\| > C} \rho(y | x) \to 0$ as $C \to \infty$.

Next, we have

$$
\sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{|u'y - M_n(u'y | x)|}{M_n(|u'y - M_n(u'y | x)| | x)} - \sup_{\|u\|=1} \frac{|u'y - M(u'y | x)|}{M(|u'y - M(u'y | x)| | x)}
\leq \sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{u'y}{M_n(|u'y - M_n(u'y | x)| | x)} - \frac{M(u'y | x)}{M(|u'y - M(u'y | x)| | x)}
+ \sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{M_n(u'y | x)}{M_n(|u'y - M_n(u'y | x)| | x)} - \frac{M(u'y | x)}{M(|u'y - M(u'y | x)| | x)}.
$$

Since $M_n(u'y | x) \xrightarrow{a.s.} M(u'y | x)$ and $M_n(|u'y - M_n(u'y | x)| | x) \xrightarrow{a.s.} M(|u'y - M(u'y | x)| | x)$ as $n \to \infty$ uniformly in $\{u \in \mathbb{R}^p | \|u\| = 1\}$, it follows that

$$
\sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{u'y}{M_n(|u'y - M_n(u'y | x)| | x)} - \frac{M(u'y | x)}{M(|u'y - M(u'y | x)| | x)} \xrightarrow{a.s.} 0,
$$

$$
\sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{M_n(u'y | x)}{M_n(|u'y - M_n(u'y | x)| | x)} - \frac{M(u'y | x)}{M(|u'y - M(u'y | x)| | x)} \xrightarrow{a.s.} 0
$$
as $n \to \infty$. Hence,

$$
\sup_{\|y\| \leq C} \sup_{\|u\|=1} \frac{|u'y - M_n(u'y | x)|}{M_n(|u'y - M_n(u'y | x)| | x)} - \sup_{\|u\|=1} \frac{|u'y - M(u'y | x)|}{M(|u'y - M(u'y | x)| | x)} \xrightarrow{a.s.} 0 \text{ as } n \to \infty.
$$
The following results are required for the proofs of Theorems 4.2, 4.3 and 4.4.

**Lemma B.3.** Suppose (4.2) holds. Then, for any \( \epsilon > 0 \), \( \delta > 0 \) and any sequence \( \{ \alpha_n \} \) with \( \alpha_n \xrightarrow{} \alpha \) as \( n \rightarrow \infty \), where \( \alpha \) is a real number, we have

\[
P[D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x)] \geq 1 - \epsilon
\]

for all sufficiently large \( n \).

**Proof.** If \( \alpha + \delta > \max_y \rho(y \mid x) \), then \( D(\alpha + \delta \mid x) = \emptyset \subseteq D_n(\alpha_n \mid x) \) with probability 1. On the other hand, if \( \alpha - \delta < \min_y \rho(y \mid x) \), then \( D(\alpha - \delta \mid x) \) is the entire response space, and as a result \( D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x) \) with probability 1. So we assume

\[
\delta \leq \min \left\{ \max_y \rho(y \mid x) - \alpha, \alpha - \min_y \rho(y \mid x) \right\}.
\]

(B.4)

Given \( \epsilon > 0 \) and \( \delta > 0 \), we take

\[
\epsilon_1 = \epsilon_2 = \frac{\epsilon}{4}, \delta_1 = \delta_2 = \frac{\delta}{4}.
\]

There exists \( N_1(\epsilon_1, \delta_1) \) such that for all \( n \geq N_1(\epsilon_1, \delta_1) \),

\[
P[\alpha - \delta_1 < \alpha_n < \alpha + \delta_1] > 1 - \epsilon_1.
\]

(B.5)

Similarly, under (4.2), there exists \( N_2(\epsilon_2, \delta_2) \) such that for all \( n \geq N_2(\epsilon_2, \delta_2) \),

\[
P[\rho(y \mid x) - \delta_2 < \rho_n(y \mid x) < \rho(y \mid x) + \delta_2 \ \forall y] > 1 - \epsilon_2.
\]

(B.6)

Using (B.4), (B.5), and (B.6), we have for all \( n \geq \max \{N_1(\epsilon_1, \delta_1), N_2(\epsilon_2, \delta_2)\} \),

\[
P[\alpha - \delta_1 < \alpha_n < \alpha + \delta_1 \text{ and } \rho(y \mid x) - \delta_2 < \rho_n(y \mid x) < \rho(y \mid x) + \delta_2 \ \forall y] > 1 - (\epsilon_1 + \epsilon_2) \Rightarrow P[\alpha_n - \delta_1 < \alpha \text{ and } \alpha + \delta_2 \leq \rho(y \mid x) - \delta_2 < \rho_n(y \mid x) \ \forall y \text{ s.t. } \rho(y \mid x) \geq \alpha + \delta] > 1 - (\epsilon_1 + \epsilon_2) \Rightarrow P[D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n + \delta - (\delta_1 + \delta_2) \mid x)] > 1 - (\epsilon_1 + \epsilon_2).
\]

(B.7)

Also, again using (B.4), (B.5), and (B.6), we have for all \( n \geq \max \{N_1(\epsilon_1, \delta_1), N_2(\epsilon_2, \delta_2)\} \),

\[
P[\alpha - \delta_1 < \alpha_n < \alpha + \delta_1 \text{ and } \rho(y \mid x) - \delta_2 < \rho_n(y \mid x) < \rho(y \mid x) + \delta_2 \ \forall y] > 1 - (\epsilon_1 + \epsilon_2) \Rightarrow P[\alpha < \alpha_n + \delta_1 \text{ and } \rho_n(y \mid x) < \rho(y \mid x) + \delta_2 < \rho_n(y \mid x) < \rho(y \mid x) + \delta_2 \ \forall y \text{ s.t. } \rho(y \mid x) < \alpha - \delta] > 1 - (\epsilon_1 + \epsilon_2) \Rightarrow P[D_n(\alpha_n - (\delta - (\delta_1 + \delta_2)) \mid x) \subseteq D(\alpha - \delta \mid x)] > 1 - (\epsilon_1 + \epsilon_2).
\]

(B.8)

So, from (B.7) and (B.8) we get that, for all \( n \geq \max \{N_1(\epsilon_1, \delta_1), N_2(\epsilon_2, \delta_2)\} \),

\[
P[D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x)] \geq P[D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n + (\delta - (\delta_1 + \delta_2)) \mid x) \text{ and } D_n(\alpha_n - (\delta - (\delta_1 + \delta_2)) \mid x) \subseteq D(\alpha - \delta \mid x)] > 1 - 2(\epsilon_1 + \epsilon_2) = 1 - \epsilon.
\]

□
Lemma B.4. Suppose \((\ref{C(iii)})\), \((\ref{4.1})\) and \((\ref{4.2})\) hold. Then, for any \(\beta\), \(\mu_n(D_n(\beta | x) | x) \xrightarrow{P} \mu(D(\beta | x) | x)\) as \(n \to \infty\).

Proof. Note that

\[
|\mu_n(D_n(\beta | x) | x) - \mu(D(\beta | x) | x)|
\leq |\mu_n(D_n(\beta | x) | x) - \mu_n(D(\beta | x) | x)| + |\mu_n(D(\beta | x) | x) - \mu(D(\beta | x) | x)|. \tag{B.9}
\]

From \((\ref{C(iii)})\) and \((\ref{4.1})\), we get that

\[
|\mu_n(D(\beta | x) | x) - \mu(D(\beta | x) | x)| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty. \tag{B.10}
\]

We shall show that \(|\mu_n(D_n(\beta | x) | x) - \mu_n(D(\beta | x) | x)| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty\).

Given any \(\epsilon, \delta > 0\), take \(\epsilon_1 = \epsilon/2\) and \(\delta_1 = \delta/2\). Since \(\mu(\{y \mid \rho(y | x) = \beta\} | x) = 0\) under \((\ref{C(iii)})\), we can find \(\delta_2 > 0\) such that

\[
\mu(\{y \mid |\beta - \delta_2 \leq \rho(y | x) < \beta + \delta_2\} | x) < \delta_1. \tag{B.11}
\]

Denote \(B(\beta, \delta_2) = \{y \mid |\beta - \delta_2 \leq \rho(y | x) < \beta + \delta_2\}\). Under \((\ref{4.1})\) and from \((\ref{B.11})\), we get that there exists \(N_1(\delta_1, \epsilon_1)\) such that for all \(n \geq N_1(\delta_1, \epsilon_1)\),

\[
P[\mu_n(B(\beta, \delta_2) | x) < 2\delta_1] \geq P[|\mu_n(B(\beta, \delta_2) | x) - \mu(B(\beta, \delta_2) | x)| < \delta_1] \geq 1 - \epsilon_1. \tag{B.12}
\]

Denote the events

\[
A_n^{(1)}(\beta, \delta_1, \delta_2) = \{\mu_n(B(\beta, \delta_2) | x) < 2\delta_1\},
A_n^{(2)}(\beta, \delta_2) = \{D(\beta + \delta_2 | x) \subseteq D_n(\beta | x) \subseteq D(\beta - \delta_2 | x)\}.
\]

From Lemma \(\ref{B.3}\) we get that there exists \(N_2(\epsilon_1, \delta_2)\) such that for all \(n \geq N_2(\epsilon_1, \delta_2)\),

\[
P[A_n^{(2)}(\beta, \delta_2)] > 1 - \epsilon_1. \tag{B.13}
\]

Since \(D(\beta + \delta_1 | x) \subseteq D(\beta | x) \subseteq D(\beta - \delta_1 | x)\), we have for \(\omega \in A_n^{(2)}(\beta, \delta_2)\),

\[
|\mu_n(D_n(\beta | x) | x) - \mu_n(D(\beta | x) | x)| \leq \mu_n(B(\beta, \delta_2) | x). \tag{B.14}
\]

So, from \((\ref{B.12})\), \((\ref{B.13})\) and \((\ref{B.14})\), we get that for all \(n \geq \max\{N_1(\delta_1, \epsilon_1), N_2(\epsilon_1, \delta_2)\}\),

\[
P[|\mu_n(D_n(\beta | x) | x) - \mu_n(D(\beta | x) | x)| < \delta]
\geq P[|\mu_n(D_n(\beta | x) | x) - \mu_n(D(\beta | x) | x)| \leq \mu_n(B(\beta, \delta_2) | x) \text{ and } \mu_n(B(\beta, \delta_2) | x) < 2\delta_1]
\geq P[A_n^{(2)}(\beta, \delta_2) \cap A_n^{(1)}(\beta, \delta_1, \delta_2)] > 1 - \epsilon. \tag{B.15}
\]

The proof follows from \((\ref{B.9})\), \((\ref{B.10})\) and \((\ref{B.15})\). \(\square\)

Lemma B.5. Under \((\ref{C(iii)})\), \((\ref{C(iv)})\) \((\ref{4.1})\) and \((\ref{4.2})\), \(\alpha_n(r) \xrightarrow{P} \alpha(r)\) as \(n \to \infty\) for any \(0 < r < 1\).

Proof. Given \(\epsilon > 0\) and \(\delta > 0\), denote

\[
\delta_1 = \frac{r - \mu(D(\alpha(r) + \delta | x) | x)}{2}, \quad \delta_2 = \frac{\mu(D(\alpha(r) - \delta | x) | x) - r}{2}.
\]

25
By definition of $\alpha(r)$, $\delta_1 > 0$. From [C(iv)] we have $\mu(\{y \mid \alpha(r) - \delta \leq \rho(y \mid x) < \alpha(r)\} \mid x) > 0$, which ensures $\delta_2 > 0$. Define the events

- $A_n(\delta) = \{\alpha_n(r) > \alpha(r) + \delta\}$,
- $B_n(\delta) = \{\alpha_n(r) < \alpha(r) - \delta\}$,
- $C_n^{(1)}(\delta_1) = \{\mu_n(D_n(\alpha(r) + \delta \mid x) \mid x) - \mu(D(\alpha(r) + \delta \mid x) \mid x) < \delta_1\}$,
- $C_n^{(2)}(\delta_2) = \{\mu_n(D_n(\alpha(r) - \delta \mid x) \mid x) - \mu(D(\alpha(r) - \delta \mid x) \mid x) < \delta_2\}$.

From Lemma B.4 we get that there exists $N(\epsilon, \delta_1, \delta_2)$ such that for all $n \geq N(\epsilon, \delta_1, \delta_2)$,

$$P[C_n^{(1)}(\delta_1)] > 1 - (\epsilon/2) \text{ and } P[C_n^{(2)}(\delta_2)] > 1 - (\epsilon/2).$$

(B.16)

Note that,

- for $\omega \in A_n(\delta)$, $\mu_n(D_n(\alpha(r) + \delta \mid x) \mid x) \geq r$,
- for $\omega \in B_n(\delta)$, $\mu_n(D_n(\alpha(r) - \delta \mid x) \mid x) < r$,
- for $\omega \in C_n^{(1)}(\delta_1)$, $\mu_n(D_n(\alpha(r) + \delta \mid x) \mid x) < \mu(D(\alpha(r) + \delta \mid x) \mid x) + \delta_1 < r$,
- for $\omega \in C_n^{(2)}(\delta_2)$, $\mu_n(D_n(\alpha(r) - \delta \mid x) \mid x) > \mu(D(\alpha(r) - \delta \mid x) \mid x) - \delta_2 > r$.

So, $A_n(\delta) \cap C_n^{(1)}(\delta_1) = \emptyset$ and $B_n(\delta) \cap C_n^{(2)}(\delta_2) = \emptyset$. Consequently, from (B.16) it follows that for all $n \geq N(\epsilon, \delta_1, \delta_2)$,

$$P[|\alpha_n(r) - \alpha(r)| > \delta] = P[A_n(\delta) \cup B_n(\delta)] \leq 1 - P[C_n^{(1)}(\delta_1) \cap C_n^{(2)}(\delta_2)] \leq \epsilon.
\]

\[\square\]

Lemma B.6. Suppose that [C(v), C(v)], and (4.2) hold. Then, for any sequence $\{\alpha_n\}$ with $\alpha_n \xrightarrow{p} \alpha$ as $n \to \infty$, $d_H(D_n(\alpha_n \mid x), D(\alpha \mid x)) \xrightarrow{p} 0$ as $n \to \infty$.

Proof. Recall that for any pair of sets $A$ and $B$, $d_H(A, B) = \max\{\sup_{u \in A} \inf_{v \in B} \|u - v\|, \sup_{v \in B} \inf_{u \in A} \|u - v\|\}$. Since $D(\alpha + \delta \mid x) \subseteq D(\alpha \mid x) \subseteq D(\alpha - \delta \mid x)$ for any $\delta > 0$, we have

$$d_H(D(\alpha + \delta \mid x), D(\alpha - \delta \mid x)) \leq d_H(D(\alpha \mid x), D(\alpha - \delta \mid x)) + d_H(D(\alpha + \delta \mid x), D(\alpha \mid x)).$$

(B.17)

Denote $d_0(u, A) = \inf\{\|u - v\| \mid v \in A\}$ for a point $u$ and a set $A$. So, $d_H(D(\alpha \mid x), D(\alpha - \delta \mid x)) = \sup\{d_0(y, D(\alpha \mid x)) \mid y \in D(\alpha - \delta \mid x)\}$. Suppose, if possible, $\lim_{\delta \to 0^+} d_H(D(\alpha \mid x), D(\alpha - \delta \mid x)) > 0$. Then, there exists a constant $\delta_1 > 0$ and a sequence $\{y_n\}$ such that $\rho(y_n \mid x) \geq \alpha(1 - 0.5(1/n))$ and $d_0(y_n, D(\alpha \mid x)) \geq \delta_1$ for all $n$. Under [C(v)] $D((\alpha/2) \mid x)$ is bounded, and so $\{y_n\}$ is a bounded sequence in $\mathbb{R}^p$, the response space. Consequently, $\{y_n\}$ has convergent subsequence $\{y_{n_k}\}$, with, say, $y_{n_k} \to y_0$ as $k \to \infty$. From [C(v)], we have $\rho(y_0 \mid x) = \lim_k \rho(y_{n_k} \mid x) \geq \alpha$ and $d_0(y_0, D(\alpha \mid x)) = \lim_k d_0(y_{n_k}, D(\alpha \mid x)) \geq \delta_1 > 0$, which contradict themselves. Therefore,

$$\lim_{\delta \to 0^+} d_H(D(\alpha \mid x), D(\alpha - \delta \mid x)) = 0. \quad \text{(B.18)}
$$

Next, we look into the term $d_H(D(\alpha + \delta \mid x), D(\alpha \mid x))$. We have $d_H(D(\alpha + \delta \mid x), D(\alpha \mid x)) = \sup\{d_0(y, D(\alpha + \delta \mid x)) \mid y \in D(\alpha \mid x)\}$. Suppose, if possible, $\lim_{\delta \to 0^+} d_H(D(\alpha + \delta \mid x), D(\alpha \mid x))$
> d_H(D_0(\alpha \mid x), D(\alpha \mid x)). Then, there are \delta_2, \delta_3 > 0 and a sequence \{y_n\} \subset D(\alpha \mid x) such that d_0(y_n, D(\alpha(1 + (1/n)) \mid x)) > \delta_2 > \delta_3 > d_H(D_0(\alpha \mid x), D(\alpha \mid x)) for all sufficiently large \(n\). From [C(v)] we have \(D(\alpha \mid x)\) closed and bounded. So, \(\{y_n\}\) has a convergent subsequence \(\{y_{n_k}\}\) with \(\lim_{k \to \infty} y_{n_k} = y_0 \in D(\alpha \mid x)\). Since \(d_0(y_0, D(\alpha(1 + (1/n_k)) \mid x)) \geq d_0(y_{n_k}, D(\alpha(1 + (1/n_k)) \mid x)) - \|y_{n_k} - y_0\|\), it follows that \(d_0(y_0, D(\alpha(1 + (1/n_k)) \mid x)) > \delta_3 > d_0(y_0, D_0(\alpha \mid x))\) for all sufficiently large \(k\). This implies that there is \(y' \in D(\alpha \mid x)\) with \(\delta_3 > ||y_0 - y'||\). Since \(\rho(y' \mid x) > \alpha\), we have \(y' \in D(\alpha(1 + (1/n_k)) \mid x)\) for all sufficiently large \(k\), but this leads to a contradiction as then \(||y_0 - y'|| \geq d_0(y_0, D(\alpha(1 + (1/n_k)) \mid x)) > \delta_3\). Hence, we have

\[
\lim_{\delta \to 0^+} d_H(D(\alpha + \delta \mid x), D(\alpha \mid x)) = d_H(D_0(\alpha \mid x), D(\alpha \mid x)). \tag{B.19}
\]

From (B.19) and [C(vi)] we get

\[
\lim_{\delta \to 0^+} d_H(D(\alpha + \delta \mid x), D(\alpha \mid x)) = 0. \tag{B.20}
\]

Therefore, from (B.17), (B.18) and (B.20), we get \(d_H(D(\alpha + \delta \mid x), D(\alpha - \delta \mid x)) \to 0\) as \(\delta \to 0^+\). Given any \(\epsilon > 0\) and \(\delta > 0\), take \(\delta_1 > 0\) such that

\[
d_H(D(\alpha + \delta_1 \mid x), D(\alpha - \delta_1 \mid x)) < \delta.
\]

Denote the event \(A_n(\delta_1) = \{D(\alpha + \delta_1 \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta_1 \mid x)\}\). From Lemma [B.3], we get that there exists \(N(\epsilon, \delta)\) such that for all \(n \geq N(\epsilon, \delta_1)\), \(P[A_n(\delta_1)] > 1 - \epsilon\). Also, for \(\omega \in A_n(\delta)\),

\[
d_H(D_n(\alpha_n \mid x), D(\alpha \mid x)) \leq d_H(D(\alpha + \delta_1 \mid x), D(\alpha - \delta_1 \mid x)) < \delta,
\]

which implies

\[
P[d_H(D_n(\alpha_n \mid x), D(\alpha \mid x)) \leq \delta] \geq P[A_n(\delta_1)] > 1 - \epsilon
\]

for all \(n \geq N(\epsilon, \delta_1)\). Hence, \(d_H(D_n(\alpha_n \mid x), D(\alpha \mid x)) \xrightarrow{P} 0\) as \(n \to \infty\). \(\Box\)

**Lemma B.7.** Let \(\{\alpha_n\}\) be a sequence of random variables with \(\alpha_n \xrightarrow{P} \alpha > 0\) as \(n \to \infty\). Suppose that \([C(iii)], [C(v)], \tag{4.1}\) and \([4.2]\) hold. Then,

\[
\int y I(y \in D_n(\alpha_n \mid x)) \mu_n(dy \mid x) \xrightarrow{P} \int y I(y \in D(\alpha \mid x)) \mu(dy \mid x) \quad \text{as } n \to \infty.
\]

**Proof.** We have

\[
\left\| \int y I(y \in D_n(\alpha_n \mid x)) \mu_n(dy \mid x) - \int y I(y \in D(\alpha \mid x)) \mu(dy \mid x) \right\| \\
\leq \left\| \int y I(y \in D_n(\alpha_n \mid x)) \mu_n(dy \mid x) - \int y I(y \in D(\alpha \mid x)) \mu_n(dy \mid x) \right\| \\
+ \left\| \int y I(y \in D(\alpha \mid x)) \mu_n(dy \mid x) - \int y I(y \in D(\alpha \mid x)) \mu(dy \mid x) \right\| \tag{B.21}
\]

Define the event \(A_n(\delta) = \{D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x)\}\), where \(\delta > 0\). Since \(\alpha > 0\), from [C(v)] we get that there is \(\delta_1 > 0\) such that for all \(0 < \delta \leq \delta_1\), \(D(\alpha - \delta \mid x)\) is a bounded set. For any \(0 < \delta \leq \delta_1\), when the event \(A_n(\delta)\) occurs, we have

\[
\left\| \int y I(y \in D_n(\alpha_n \mid x)) \mu_n(dy \mid x) - \int y I(y \in D(\alpha \mid x)) \mu_n(dy \mid x) \right\|
\]

27
Lemma B.8. Suppose (4.2) and C(iii) hold, and \( \mu(\cdot | x) \) has a continuous density on \( \mathbb{R}^p \), which is everywhere positive. Then, for any sequence \( \{\alpha_n\} \) with \( \alpha_n \xrightarrow{p} \alpha > 0 \) as \( n \to \infty \), 
\[
\lambda(D_n(\alpha_n | x)) \xrightarrow{p} \lambda(D(\alpha | x)) \quad \text{as} \quad n \to \infty,
\]
where \( \lambda(\cdot) \) is the Lebesgue measure on \( \mathbb{R}^p \).
Proof. Given any $\epsilon, \delta > 0$, from Lemma [B.3], we have
\[
P[D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x)] > 1 - \epsilon \tag{B.27}
\]
for all sufficiently large $n$. Define the event
\[
A_n(\delta) = \{D(\alpha + \delta \mid x) \subseteq D_n(\alpha_n \mid x) \subseteq D(\alpha - \delta \mid x)\}.
\]
When the event $A_n(\delta)$ occurs, we have
\[
|\lambda(D_n(\alpha_n \mid x)) - \lambda(D(\alpha \mid x))| \leq \lambda(\{y \mid \alpha - \delta \leq \rho(y \mid x) < \alpha + \delta\}). \tag{B.28}
\]
From the assumptions in the lemma and using the Radon-Nikodym theorem, we get that $\lambda(\cdot)$ has a Radon-Nikodym derivative $f(\cdot)$ with respect to $\mu(\cdot \mid x)$, which is continuous and bounded over bounded sets. From [C(iii)], we have $\mu(\{y \mid \alpha - \delta \leq \rho(y \mid x) < \alpha + \delta\} \mid x) \to 0$ as $\delta \to 0^+$. So,
\[
\lambda(\{y \mid \alpha - \delta \leq \rho(y \mid x) < \alpha + \delta\}) \to 0 \text{ as } \delta \to 0^+. \tag{B.29}
\]
Therefore, from (B.27), (B.28) and (B.29), we have $\lambda(D_n(\alpha_n \mid x)) \xrightarrow{P} \lambda(D(\alpha \mid x))$ as $n \to \infty$. \qed

Lemma B.9. For any pair of sets $A$ and $B$, $|\text{Diameter}(A) - \text{Diameter}(B)| \leq 2d_H(A, B)$.

Proof. Recall that $d_H(A, B) = \inf\{\epsilon \mid A \subseteq B^\epsilon, B \subseteq A^\epsilon\}$, where $A^\epsilon$ and $B^\epsilon$ denote the $\epsilon$-neighborhoods of $A$ and $B$, respectively. Take any $u, v \in A$. From the definition of $d_H(A, B)$, we get that given any $\epsilon > 0$, there are $u', v' \in B$ such that
\[
\|u - u'\| < d_H(A, B) + \epsilon \quad \text{and} \quad \|v - v'\| < d_H(A, B) + \epsilon.
\]
Since $\|u' - v'\| \leq \text{Diameter}(B)$, we have
\[
\|u - v\| \leq \|u - u'\| + \|v - v'\| + \|u' - v'\| \leq 2d_H(A, B) + 2\epsilon + \text{Diameter}(B).
\]
Since $u, v$ are arbitrary points in $A$, we have
\[
\text{Diameter}(A) - \text{Diameter}(B) \leq 2d_H(A, B) + 2\epsilon. \tag{B.30}
\]
Similarly, we can show that
\[
\text{Diameter}(B) - \text{Diameter}(A) \leq 2d_H(A, B) + 2\epsilon. \tag{B.31}
\]
Since $\epsilon > 0$ is arbitrary, from (B.30) and (B.31), we have $|\text{Diameter}(A) - \text{Diameter}(B)| \leq 2d_H(A, B)$. \qed

Lemma B.10. Let $\{\alpha_n\}$ be a sequence of random variables with $\alpha_n \xrightarrow{P} \alpha$ as $n \to \infty$. Define $D_n'(\alpha_n \mid x) = \{Y_i \mid \rho_n(Y_i \mid x) \geq \alpha_n; i = 1, ..., n\}$. Suppose that for any open set $G$ with $G \cap D(\alpha \mid x) \neq \emptyset$, we have $P[Y \in G \cap D(\alpha \mid x)] > 0$. Then, under $[C(\nu)], [C'(\nu)]$ and (4.2), $d_H(D_n'(\alpha_n \mid x), D(\alpha \mid x)) \xrightarrow{P} 0$ as $n \to \infty$. 

29
Proof. Note that
\[
d_H(D'_n(\alpha_n | x), D(\alpha | x)) = \max \left\{ \sup_{u \in D'_n(\alpha_n | x)} d_0(u, D(\alpha | x)), \sup_{v \in D(\alpha | x)} d_0(v, D'_n(\alpha_n | x)) \right\}.
\]
(B.32)

Now, given any \(\epsilon, \delta > 0\), it follows from Lemma B.3 that \(P[D'_n(\alpha_n | x) \subset D_n(\alpha_n | x) \subseteq D(\alpha - \delta | x)] > 1 - \epsilon\) for all sufficiently large \(n\). Denote the event \(A_n(\delta) = \{D'_n(\alpha_n | x) \subseteq D(\alpha - \delta | x)\}\). So, for \(\omega \in A_n(\delta)\), sup \(d_0(u, D(\alpha | x)) | u \in D'_n(\alpha_n | x)\) \(\leq\) sup \(d_0(u, D(\alpha | x)) | u \in D(\alpha - \delta | x)\). Since sup \(d_0(u, D(\alpha | x)) | u \in D(\alpha - \delta | x)\) = \(d_H(D(\alpha | x), D(\alpha - \delta | x))\), from (B.18) it follows that
\[
\sup_{u \in D'_n(\alpha_n | x)} d_0(u, D(\alpha | x)) \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty.
\]
(B.33)

Next, we consider the term sup \(d_0(v, D'_n(\alpha_n | x)) | v \in D(\alpha | x)\). Let \(\epsilon, \delta > 0\) be any given numbers. From C(v), we get that \(D(\alpha | x)\) is a compact subset of \(\mathbb{R}^p\). So, we can cover \(D(\alpha | x)\) by a finite number of pairwise disjoint semi-open hypercubes of long-diagonal length \((\delta/2)\), such that the interior of each hypercube has a non-empty intersection with \(D(\alpha | x)\). Let the number of such hypercubes be \(n_1\) and the collection of hypercubes be denoted as \(\{C_1, \cdots, C_{n_1}\}\). So, under the assumption of the lemma, we have
\[
\min\{P[Y \in C_i \cap D(\alpha | x)] | i = 1, \cdots, n_1\} \\
\geq \min\{P[Y \in \text{int}(C_i) \cap D(\alpha | x)] | i = 1, \cdots, n_1\} > 0,
\]
where \(\text{int}(C)\) denotes the interior of a set \(C\). Consider the event
\[
A_n = \{\text{Each } (C_i \cap D(\alpha | x)) \text{ contains at least one observation } Y_j, j = 1, \cdots, n\}.
\]

Since the hypercubes are pairwise disjoint, from (B.34), we have \(P[A_n] \to 1\) as \(n \to \infty\). Let \(n_2\) be an integer such that for all \(n \geq n_2\),
\[
P[A_n] > 1 - (\epsilon/2).
\]
(B.35)

From (B.20), we can find \(\delta_1 > 0\) sufficiently small such that
\[
\sup\{d_0(v, D(\alpha + 2\delta_1 | x)) | v \in D(\alpha | x)\} < (\delta/2).
\]
(B.36)

So, using the triangle inequality and (B.36), we have
\[
\sup\{d_0(v, D'_n(\alpha_n | x)) | v \in D(\alpha | x)\} \\
\leq \sup\{d_0(v, D(\alpha + 2\delta_1 | x)) | v \in D(\alpha | x)\} + \sup\{d_0(v, D'_n(\alpha_n | x)) | v \in D(\alpha + 2\delta_1 | x)\} \\
< (\delta/2) + \sup\{d_0(v, D'_n(\alpha_n | x)) | v \in D(\alpha + 2\delta_1 | x)\}.
\]
(B.37)

Define the event
\[
B_n = \left\{ |\alpha_n - \alpha| < \delta_1 \text{ and } \sup_{y \in \mathbb{R}^p} |\rho_n(y | x) - \rho(y | x)| < \delta_1 \right\}.
\]
Since $\alpha_n \xrightarrow{P} \alpha$ as $n \to \infty$, and (4.2) is assumed to be satisfied, there is an integer $n_3$ such that for all $n \geq n_3$,

$$P[B_n] > 1 - (\epsilon/2). \quad (B.38)$$

When the event $B_n$ occurs, we have $\rho_n(Y_j | x) \geq \alpha_n$ for any sample observation $Y_j \in D(\alpha + 2\delta_1 | x)$. Consequently, when the event $(A_n \cap B_n)$ occurs, we have

$$\sup\{d_0(v, D_n' (\alpha_n | x)) \mid v \in D(\alpha + 2\delta_1 | x)\} < (\delta/2) \quad (B.39)$$

since $D(\alpha + 2\delta_1 | x) \subset D(\alpha | x)$. Therefore, from (B.35), (B.37), (B.38) and (B.39), we have for all $n \geq \max\{n_2, n_3\}$,

$$P \left[ \sup_{v \in D(\alpha | x)} d_0(v, D_n' (\alpha_n | x)) \mid v \in D(\alpha | x) \right] < \delta \quad (B.40)$$

implies that

$$\sup_{v \in D(\alpha | x)} d_0(v, D_n' (\alpha_n | x)) \xrightarrow{P} 0 \quad \text{as } n \to \infty. \quad (B.41)$$

The proof is complete from (B.32), (B.33) and (B.41). \hfill \Box

References

Arcones, M. A. and Gine, E. (1993). Limit theorems for u-processes. *The Annals of Probability*, 21(3):1494–1542.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Chen, Y., Dang, X., Peng, H., and Bart Jr, H. L. (2009). Outlier detection with the kernelized spatial depth function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):288–305.

Chowdhury, J. and Chaudhuri, P. (2016). Nonparametric depth and quantile regression for functional data. *arXiv preprint arXiv:1607.03752* (to appear in *Bernoulli*).

Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of statistical planning and inference*, 103(1):311–329.

Dette, H. and Marchlewski, M. (2010). A robust test for homoscedasticity in nonparametric regression. *Journal of Nonparametric Statistics*, 22(6):723–736.

Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *Journal of the Royal Statistical Society: Series B*, 60(4):693–708.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827.
Dutta, S. and Ghosh, A. K. (2012). On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics*, 64(3):657–676.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Number 66 in Monographs on Statistics and Applied Probability. CRC Press.

Ghosh, A. K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics*, 32(2):327–350.

Gupta, A. and Tang, J. (1984). Distribution of likelihood ratio statistic for testing equality of covariance matrices of multivariate Gaussian models. *Biometrika*, 71(3):555–559.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.

He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *The Annals of Statistics*, 25(2):495–504.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30.

Holgersson, H. and Shukur, G. (2004). Testing for multivariate heteroscedasticity. *Journal of Statistical Computation and Simulation*, 74(12):879–896.

Jörnsten, R. (2004). Clustering and classification based on the $l_1$ data depth. *Journal of Multivariate Analysis*, 90(1):67–89.

Koenker, R. (2005). *Quantile Regression*. Number 38 in Econometric Society Monographs. Cambridge University Press.

Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25(2):435–477.

Li, J., Cuesta-Albertos, J. A., and Liu, R. Y. (2012). Dd-classifier: Nonparametric classification procedure based on dd-plot. *Journal of the American Statistical Association*, 107(498):737–753.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414.

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion and a rejoinder by Liu and Singh). *The Annals of Statistics*, 27(3):783–858.

Massé, J.-C. (2009). Multivariate trimmed means based on the Tukey depth. *Journal of Statistical Planning and Inference*, 139(2):366–384.

Munkres, J. R. (2000). *Topology*. Prentice Hall, 2 edition.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer Science & Business Media.
Rousseeuw, P. J. and Ruts, I. (1996). Algorithm as 307: Bivariate location depth. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4):516–526.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: a bivariate boxplot. *The American Statistician*, 53(4):382–387.

Rousseeuw, P. J. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8(3):193–203.

Serfling, R. (2002). A depth function and a scale curve based on spatial quantiles. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods*, pages 25–38. Birkhäuser Basel, Basel.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, volume 162. John Wiley & Sons.

Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley Series in Behavioral Science: Quantitative Methods. Addison-Wesley, Reading, Mass.

Van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, corrected edition.

Vardi, Y. and Zhang, C.-H. (2000). The multivariate l_1-median and associated data depth. *Proceedings of the National Academy of Sciences*, 97(4):1423–1426.

Wand, M. P. and Jones, M. C. (1994). *Kernel smoothing*. Crc Press.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372.

Zuo, Y. (2006). Multidimensional trimming based on projection depth. *The Annals of Statistics*, 34(5):2211–2251.

Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.

Zuo, Y. and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics*, 28(2):483–499.