University of Alberta

Library Release Form

Name of Author: Linglong Kong

Title of Thesis: On Multivariate Quantile Regression: Directional Approach and Application with Growth Charts

Degree: Doctor of Philosophy

Year this Degree Granted: 2009

. . . . . .
Linglong Kong

Date: . . . . . .

University of Alberta

ON MULTIVARIATE QUANTILE REGRESSION: DIRECTIONAL APPROACH AND APPLICATION WITH GROWTH CHARTS

by

Linglong Kong

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

in

Statistics

Department of Mathematical and Statistical Sciences

Edmonton, Alberta
Fall 2009

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **On Multivariate Quantile Regression: Directional Approach and Application with Growth Charts** submitted by Linglong Kong in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in *Statistics*.

. . . . . .
Dr. Byron Schmuland (Chair)

. . . . . .
Dr. Ivan Mizera (Supervisor)

. . . . . .
Dr. Narasimha Prasad

. . . . . .
Dr. Douglas Wiens

. . . . . .
Dr. Ying Cui (Educational Psychology)

. . . . . .
Dr. Ying Wei (External Examiner)

Date: . . . . . .

To my parents

Xiangyu Kong and Shuzeng Ma

and my wife

Bei Jiang

Abstract

In this thesis, we introduce a concept of directional quantile envelopes, the intersection of the halfspaces determined by directional quantiles, and show that they allow for explicit probabilistic interpretation, compared to other multivariate quantile concepts. Directional quantile envelopes provide a way to perform multivariate quantile regression: to "regress contours" on covariates. We also develop theory and algorithms for an important application of multivariate quantile regression in biometry: bivariate growth charts.

We prove that directional quantiles are continuous and derive their closed-form expression for elliptically symmetric distributions. We provide probabilistic interpretations of directional quantile envelopes and establish that directional quantile envelopes are essentially halfspace depth contours. We show that distributions with smooth directional quantile envelopes are uniquely determined by their envelopes.

We describe an estimation scheme of directional quantile envelopes and prove its affine equivariance. We establish the consistency of the estimates of directional quantile envelopes and describe their accuracy. The results are applied to estimation of bivariate extreme quantiles. One of the main contributions of this thesis is the construction of bivariate growth charts, an important application of multivariate quantile regression.

We discuss the computation of our multivariate quantile regression by developing a fast elimination algorithm. The algorithm constructs the set of active halfspaces to form a directional quantile envelope. Applying this algorithm to a large number of quantile

halfspaces, we can construct an arbitrary exact approximation of the direction quantile envelope.

In the remainder of the thesis, we exhibit the connection between depth contours and directional regression quantiles (Laine, 2001), stated without proof in Koenker (2005). Our proof uses the duality theory of primal-dual linear programming. Aiming at interpreting halfspace depth contours, we explore their properties for empirical distributions, absolutely continuous distributions and certain general distributions.

Finally, we propose a generalized quantile concept, depth quantile, inspired by halfspace depth (Tukey, 1975) and regression depth (Rousseeuw and Hubert, 1999). We study its properties in various data-analytic situations: multivariate and univariate locations, regression with and without intercept. In the end, we show an example that while the quantile regression of Koenker and Bassett (1978) fails, our concept provides sensible answers.

ACKNOWLEDGEMENTS

I would like to thank all of those who have supported and helped me throughout my Ph. D. studies.

I am deeply grateful inspirational instruction, tremendous support and invaluable guidance of my supervisor Dr. Ivan Mizera. He introduced me to quantile regression and helped find the research topic of bivariate growth charts. This dissertation has benefited from his insights and intellectual acumen. I am privileged to have worked under his supervision. I am also sincerely grateful to Dr. Ivan Mizera for being my teaching mentor and his teaching advice during my doctoral program. I believe that the research and teaching experience I have gained during my study with him has had a profound influence and impact upon my graduate education and my academic career.

My sincere thanks are extended to other members of my examining committee: Dr. Byron Schmuland, Dr. Douglas Wiens, Dr. N.G.Narasimha Prasad, Dr. Ying Cui, and Dr. Ying Wei for their thorough review, insightful comments, and helpful suggestions. I also greatly appreciate Dr. Weis complimentary comments on my thesis work and Dr. Schmuland's editorial corrections.

I also would like to thank Ms. Tara Schuetz for facilitating my doctoral study, and my friend- Ms. Anne-Marie Adachi for helping edition on my thesis.

My profound thankfulness and love to my wife, Bei Jiang, for her companionship, understanding and encouragement while I was working to complete this dissertation. I am grateful for the support in various forms from my family and friends, to mention a few: my parents-Mr. Xiangyu Kong and Mrs. Shuzeng Ma, my old brother-Mr. Lingtai

Kong, my old sister-Mrs. Qiuli Kong.

This work and the conference travels during my study are supported by Dr. Ivan Mizera, Department of Mathematical and Statistical Department, Statistical Society of Canada, and Faculty of Graduate Studies and Research. I am thankful.

Contents

List of Figures

Chapter 1

Introduction and overview of the thesis

1.1 Statistical methods using quantiles

The concept of quantile function of a univariate probability distribution is known well beyond any need of exposition; a casual reference like Shorack (2000) or the encyclopedic entry of Eubank (1986) results in the following.

Definition 1.1.1 *For $0 < p < 1$, the p-th* **quantile** *of a distribution P is defined to be*

$$Q(p) = \inf\{x : F(x) \geq p\},$$

where $F(x) = P((-\infty, x])$ is the cumulative distribution function of P.

The most prominent quantile is the one corresponding to $p = 1/2$ and is called *median*; its sample version is a robust alternative of sample mean, an estimator of population location parameter, with a breakdown point $1/2$. Other quantiles do often have traditional names too—for instance, those indexed by $p = 1/n, 2/n, \ldots, (n-1)/n$, with $n = 3, 4, 5, 9, 10$ are known as *terciles*, *quartiles*, *quintiles*, *noniles* and *deciles*, respectively; the relevant linguistic aspects are discussed by Aronson (2001). A preference for percentages in the general populace is reflected by a synonymous term *percentile*, indexed by $100p$ rather than p; we will use this form a lot in the thesis.

Quantile functions are used in both statistical applications and Monte Carlo simulations. Monte Carlo simulations employ quantile functions to produce non-uniform random or pseudorandom numbers for use in diverse types of simulation calculations. Those simulation methods are particularly useful in finance, see Jackel (2002). However, in the thesis, we will mainly focus on the statistical applications of quantiles. Quantiles play a fundamental and multifaceted role in statistics, especially in the aspect of applications. For instance, we can use Q-Q plot to compare two distributions to see how close they are; we can use Quantile-Quantile function to identify outliers; we can use quantiles to construct robust L-estimates of location; and so forth. For more information, see Eubank (1986), Gilchrist (2000), Parzen (2004) and references there. In spite of all these applications, in the thesis we intentionally ignore all the potential variety and accent only the direct probabilistic interpretations, those coming as an immediate consequence of the definition.

We will illustrate two graphic applications to make the thrust of this intent more clear. The data we are using is the weight and height of 4291 Nepali children, aged between 3 and 60 months—the data constituting a part of the Nepal Nutrition Intervention Project-Sarlahi (NNIPS, principal investigator Keith P. West, Jr., funded by the Agency of International Development). The first example is a popular tool of the direct probabilistic interpretations, boxplot. Boxplot, first introduced by Tukey (1975) and also known as box-and-whisker plot, is a convenient way to graphically explore univariate data through the five-number summaries (the smallest observation, lower quartile, median, upper quartile, and largest observation). Boxplot can be used to display differences between populations without making any assumptions of the underlying statistical distribution. The spacings between the different parts of the box help indicate the degree of dispersion and skewness in the data, and identify outliers. Boxplot is a

very useful application of univariate quantiles, which becomes almost a daily tool when visually checking the data profile. The boxplots of Nepali children weight and height are showed in the left panel of Figure 1.1, where a rough idea of the distributions of height and weight can be observed as well as one outlier in each plot.

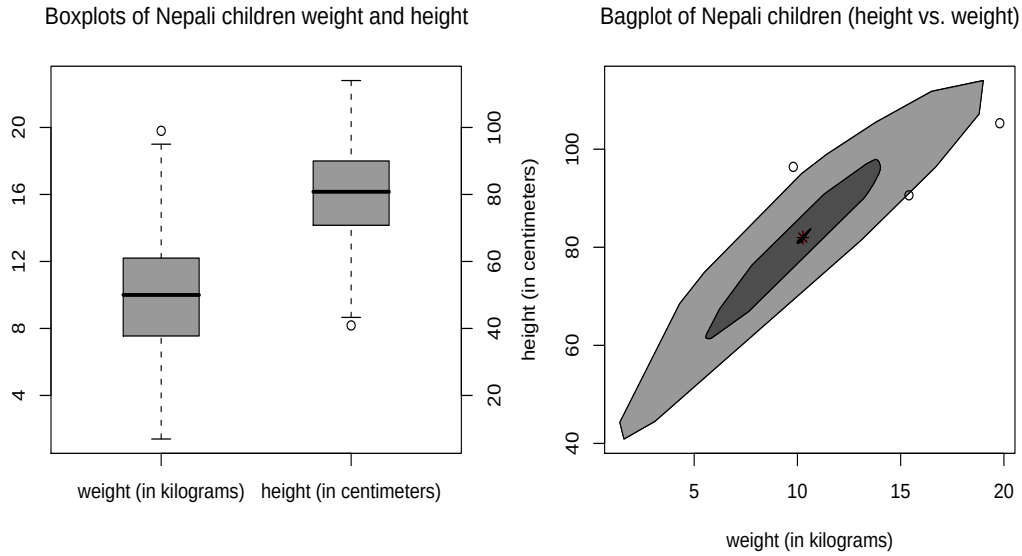

Figure 1.1: *The left panel is the boxplots of Nepali children weight and height. The right panel is the bagplot of nepali children weight and height.*

The second example is bagplot (Rousseeuw, Ruts and Tukey, 1999), a bivariate generalization of boxplot based on halfspace depth. In the bagplot, the center is the Tukey median, the point(s) with maximum halfspace depth. In the bivariate case the box of the boxplot changes to a convex hull, the bag of bagplot. In the bag are 50% of all points. The fence separates points inside the fence from points outside. The loop is defined as the convex polygon containing all points inside the fence. Observations outside the fence are outliers as they are too far away from the data's central bulk in term of their depth. Like boxplot, bagplot also visualizes several characteristics of the data:

its location (the Tukey median), spread (the size of the bag), correlation (the orientation of the bag), skewness (the shape of the bag and the loop), and tails (the points near the boundary of the loop and the outliers). The right panel of Figure 1.1 is the bagplot of Nepali children's height versus weight using the `R` package `aplpack`, implemented by Wolf (2007) based on the algorithm first proposed by Rousseeuw, Ruts and Tukey (1999). The narrow ladder shape of the bag and loop implies a linear relationship between weight and height. Three outliers are also identified by the bagplot.

If we back track, we can find that the descriptive potential of the quantiles was already pointed out by Quetelet and endorsed by Edgeworth (1886, 1893) and Galton (1888-1889). We neither deny, neglect, nor give up on more sophisticated applications; however, we strongly believe that none of these is worth losing the descriptive grip illustrated in our parable above.

1.2 Growth charts

Growth charts, also known as reference centile charts, first were conceived by Quetelet in the 19th century, and have been widely used to screen the measurements from an individual subject in the context of population values. A typical growth chart consists of a family of smooth curves representing a few selected quantiles of the distribution of some physical measurements (for instance, height, weight, head circumferences and so on) of the reference population as a function of age. When a measurement is extreme on the chart, the subject is often identified for further investigation. An extreme measurement is likely to be a reflection of some unusual underlying physical condition or maybe even some disease. Growth charts are commonly used by pediatricians, family physicians, nurses, and nutritionists to monitor peoples growth. Figure 1.2 is the latest pediatric growth charts for the weight and height of boys aged from 0 to 36 months

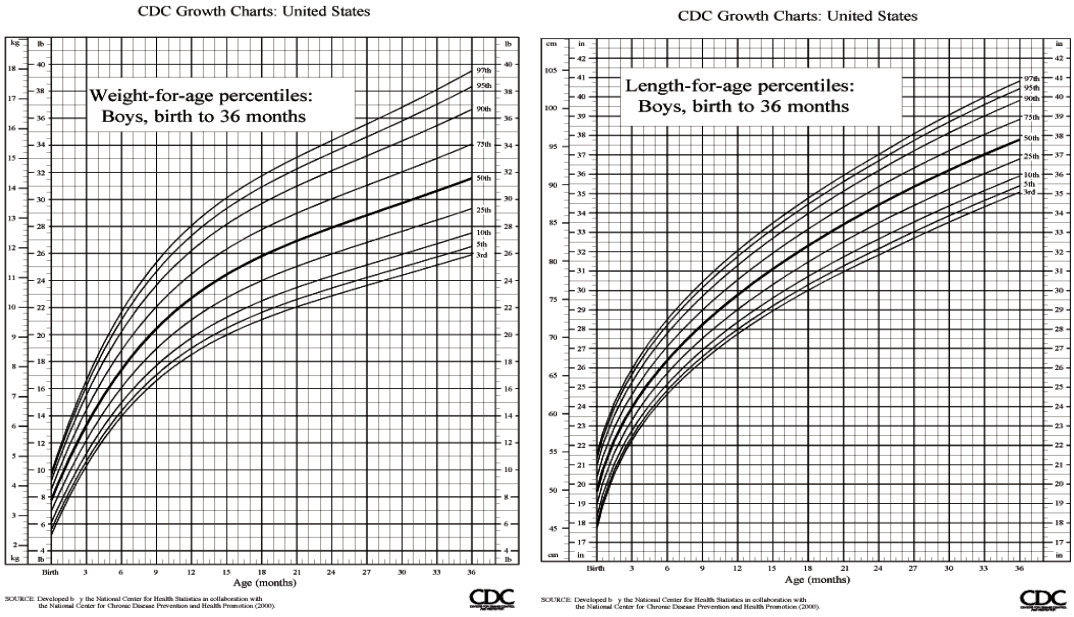

Figure 1.2: *The left panel is the weight-for-age growth charts of U.S. boys aged from* 0 *to* 36 *months. The right panel is the height-for-age growth charts of U.S. boys aged from* 0 *to* 36 *months. In both panels, the percentiles are* $\{3, 5, 10, 25, 50, 75, 90, 95, 97\}\%$.

issued by the U.S. Center for Disease Control and Prevention (May, 2000). Growth charts for the weight and height of girls with the same age range, boys and girls with different age range can be found on the website of the U.S. Center for Disease Control and Prevention, http://www.cdc.gov/.

To better understand growth charts and demonstrate how to use them to depict individuals, we plot the height and weight growth charts of Nepali children, shown in Figure 1.3, where the left panel shows the growth charts of Nepali children weight and the right panel shows the growth charts of Nepali children height with $p = .97, .95, .90, .75, .25, .10, .05, .03$. In both panels, the growth charts were regressed on the covariate, age. From the growth charts, a physician can easily figure out whether a child's height and weight measurements are "normal" and whether he or she is developing on track, or whether there are any health problems affecting the child's

growth. For instance, in the left panel the weight of subject 3110, who is at 21 months of age, is a little lower than 5% of her peers but much higher than more than 95% of her peers. While in the right panel, the height of subject 3110 is only higher than 25% of her peers but lower than 75% of her peers. Her weight and height would be described as "normal" if individually checking, but considering her low height position and relative high weight position among her peers, she may be a little overweight.

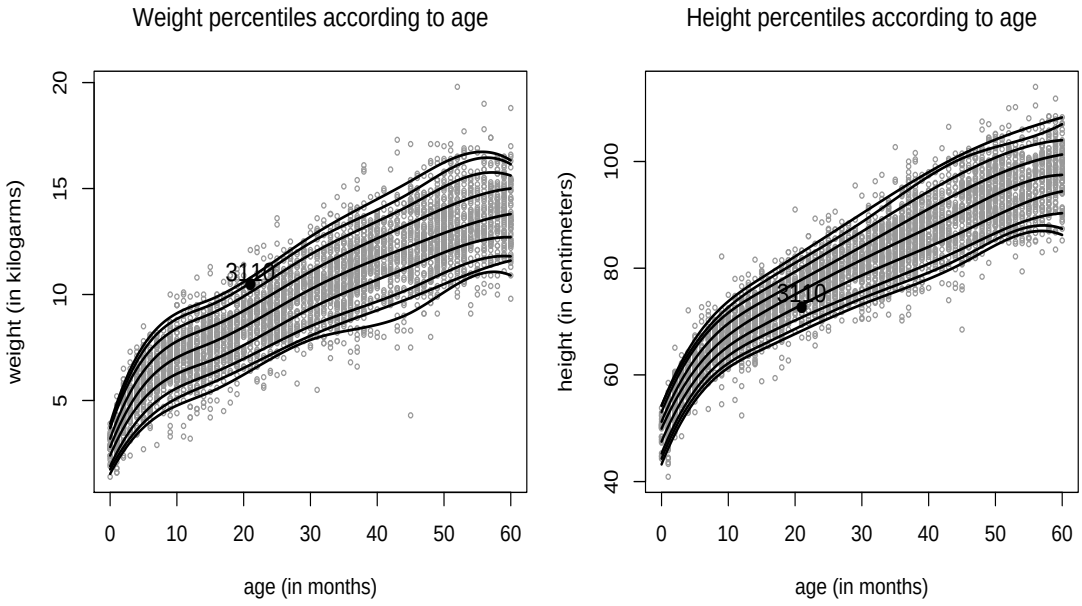

Figure 1.3: *The left panel is the weight-for-age percentile curves, weight growth charts, of Nepali children. The right panel is the height-for-age percentile curves, height growth charts, of Nepali children. The quantiles are* $\{.03, .05, .10, .25.50, .75, .90, .95, .97\}$.

The conventional method of constructing growth charts is to get the empirical quantiles of the measurements at a series of time points, and then fit a smooth polynomial curve using the empirical quantiles, see Hamill *et al.* (1979). In recent years, a number of different methods have been developed in the medical statistics literature, see Wei and He (2006) for a brief review. Most of the authors constructed growth charts based on the assumption that heights, weights, head circumferences, and other

similar measurements are normally distributed. Although adult height in reasonably homogeneous populations are known to be fairly close to normal, children's height is not normally distributed. Weights and other physical measurements are potentially even more problematic. An age-specific transformation to normality is generally used, where the Box-Cox transformation based proposals are one of the most popular choices.

One of the most successful proposals was the L-M-S method of Cole (1988). The assumption of the L-M-S method is that at any time t the measurement y can be transferred to be approximately normal by

$$z = \frac{(y/\mu)^{\lambda} - 1}{\lambda\sigma}, \ \lambda \neq 0$$
$$z = \frac{ln(y/\mu)}{\sigma}, \quad \lambda = 0,$$

where λ is the best fitting Box-Cox power for y, μ is the generalized mean and σ is the scale of the transferred variable. The estimates of power λ, the generalized mean μ and the coefficient of variation σ plotted again age t are fitted by smooth curves $L(t)$, M(t) and S(t) respectively, which together define a smooth curve for the p-th quantile $Q_t(p)$ at time t by

$$Q_t(p) = M(t)[1 + L(t)S(t)\Phi^{-1}(p)]^{1/L(t)}, \ L(t) \neq 0$$
$$Q_t(p) = M(t)exp[S(t)\Phi^{-1}(p)], \qquad L(t) = 0,$$

where $\Phi^{-1}(p)$ denotes the inverse of the standard normal distribution function. The difficulty of fitting a smooth centile curve lies in deciding whether a bump of dip observed on it at a particular age is a real feature of the data, or whether it is simply sampling error. The L-M-S method went to some way towards avoiding this problem by fitting three essentially uncorrelated curves, the $L(t)$, $M(t)$ and $S(t)$ curves, which represent the skewness, the median and the coefficient of variation of the distribution of the measurement. An improved approach to the L-M-S method, which removes one arbitrary element from the fitting process and focuses on the smoothing, the penalized log-likelihood approach, was derived by Cole and Green (1992). The main idea of the

approach is using penalized likelihood to fit the three curves as cubic splines by non-linear regression and hence the extent of smoothing required can be expressed in terms of smoothing parameters or equivalent degrees of freedom.

Two issues may arise about the transformation-based L-M-S method. One is the ability of the Box-Cox transformation to achieve the desired normality, especially over the full range of relative ages. Pere (2000) reinvestigated the Box-Cox transformation and pointed out that overall the transformation performs well but might have some issues at extreme quantiles. The other one, as demonstrated by Wei, Pere, Koenker and He (2006), is that the transformation approach may obscure features that can be unveiled by a less stringent global hypothesis methodology. In addition, interpreting the model constructed by the transformed variables is not quite clear.

To attempt solving the issues, it would be desirable to consider more principled nonparametric methods instead of any transformation. One way to accomplish the objective, as suggested by Cox and Jones in the discussion of Cole (1988) and studied by Wei, Pere, Koenker and He (2006), would be to estimate a family of conditional quantile functions by solving nonparametric quantile regression problem of the form

$$\min_{g \in \mathcal{G}} \sum_{i=1}^{n} \rho_p(y - g(t_i)), \tag{1.1}$$

subject to a smoothness requirement on the domain of the candidate functions, $\mathcal{G}$. Here the function ρ_p denotes the "check" function, that is, $\rho_p(x) = x(p - I(x < 0))$ for any $p \in (0, 1)$. Quantile regression (1.1) is a nonparametric extension of the liner quantile regression, introduced by Koenker and Bassett (1978), where the conditional quantile function g is a linear function of age t instead of a general smooth function $g \in \mathcal{G}$.

To minimize (1.1), different smoothness methods may be used to estimate g, for instance, penalty methods and regression splines. For growth charts it is convenient to parameterize the conditional quantile functions as linear combinations of a few basis

functions. B-splines are particularly convenient for this purpose. Given a choice of knots for the B-splines, estimation of the growth charts is a straightforward exercise in parametric linear regression. Solutions to such problems are linear programs and can be computed efficiently, even for very large data sets (Koenker, 2005). In Wei, Pere, Koenker and He (2006), the comparison of Cole's L-M-S method and Wei's quantile regression method based on (1.1), has been performed on the Finland modern national growth charts data set (Pere, 2000). The L-M-S method assumes a unimodal form for the conditional density, while the quantile regression method can detect bimodal conditional density which happens in Finland data. Besides that, quantile regression is capable to deal with longitudinal data with irregular time intervals. To illustrate the quantile regression method, we use the data set of Nepali children weight and height and the quantile regression method as implemented in the `R` package `quantreg`, of Koenker (2007) using a linear (B-spline) representation of the curves , to obtain the growth charts which are shown in Figure 1.3.

1.3 Bivariate growth charts

Since the conventional growth charts are usually developed from cross-sectional data, they are most useful for examining a subject with one measurement at a specific time. However, height and weight are changing over time. It is desirable to study a subject's growth path along a time period instead of at a specific time, which has been studied by using longitudinal data, for instance Cole (1994), Wei, Pere, Koenker and He (2006) and so on. Authors also considered adding in other covariates in addition to time to obtain more precise growth pathes; these covariates include past growth history, height and weight, and parental characteristics, which significantly impact the subject's growth path and partially contribute to the so called "catch up" growth phenomenon (Cole,

1994). For more details on the proposed methods, see Kim (2007), Cai and Xu (2009), and Wei and He (2006).

Adding time and other covariates into growth charts is a well developed method and has been well received in recent years. However, it remains challenging to incorporate the time (or other covariates) effects into multivariate quantiles (multivariate growth charts), for instance, the joint growth charts of height and weight. One reason to look at height and weight jointly is to have a complete picture of the subject and make accurate inference. As shown in Figure 1.3, the weight and height of subject 3110 would be described as "normal" if individually checking, but considering her low height position and relative high weight position among her peers, she may be a little overweight. Researchers may argue that we do not need height and weight bivariate growth charts; instead we can have *BMI* (Body Mass Index, ratio of weight to height square), or *ROI* (Rohrer Index, ratio of weight to height cube) growth charts. In reality, pediatricians do not just rely on these indices, they may prefer other combinations or even better look up on height and weight together. In this sense, height and weight bivariate growth charts indeed fulfill the demand.

Wei (2008) pointed out that most existing methods for constructing multivariate quantiles have difficulty incorporating the effects of covariates without strict distributional assumptions. Therefore, to construct multivariate growth charts, we first have to propose a suitable multivariate quantile, which allows the easy incorporation of covariates and also provides meaningful interpretation. As expected, Wei (2008) proposed a new form of multivariate quantiles, which is probabilistically interpretable and is flexible to capture a wide range of distributions.

For a d-dimensional random variable X, we choose μ as a component location parameter of X and $S = diag(s_i)$, $i = 1, 2, \cdots, d$, as the $d \times d$ diagonal matrix with

componentwise scale parameters of X. Let $Z = S^{-1}(X - \mu)$ be the standardized measurement and $s \in \mathbb{S}^{d-1} = \{s : \ \|s\| = 1, \text{ and } s \in \mathbb{R}^d\}$ is a standardized direction parameter. For any $p \in (0, 1)$, given a unit direction s the p-th directional quantile interval of X, proposed by wei (2008), is the closed interval $[l_p(s), u_p(s)]$ along the direction s, which contains all the points $x \in \{ks : \ k \in \mathbb{R}\}$ satisfying

$$P\left(s^T Z \leq s^T x | Z \in \{ks : \ k \in \mathbb{R}\}\right) \leq (1 + p)/2$$

and

$$P\left(s^T Z \geq s^T x | Z \in \{ks : \ k \in \mathbb{R}\}\right) \geq (1 - p)/2.$$

The p-th quantile contour $Q(p)$ is the boundary of the central set

$$R(p) = \left\{ Sx + \mu : \ x \in \bigcup_{s \in \mathbb{S}^{d-1}} [l_p(s), u_p(s)] \right\}, \tag{1.2}$$

that is, $Q(p) = \partial R(p)$. Intuitively speaking, the central set $R(p)$ consists of a family of directional quantile intervals, each of which reflects the outlyingness of measurements from μ along one spatial direction. When $d = 1$, $R(p)$ is the central interval $[F^{-1}((1 - p)/2), F^{-1}((1 + p)/2)]$ around $F^{-1}(1/2)$ with coverage p, and $Q(p)$ is its two endpoints. In general, the spatial direction s is equivalent to the constant ratio between the standardized measurements, which may reflect certain characteristics of the measurements. Using human height and weight as examples, the ration between standardized weight and height reference interval is the range of normal height and weight values among those who have a similar body composition.

In practice, the sample sizes at all time points are usually not large enough to directly calculate smooth quantile contours. Wei (2008) proposed a two-stage sampling method using a stratified quantile model to simulate the population and obtain a large enough sample to estimate the quantile contours. The constructed bivariate (height and weight) age-dependent growth charts gave pediatricians a more comprehensive

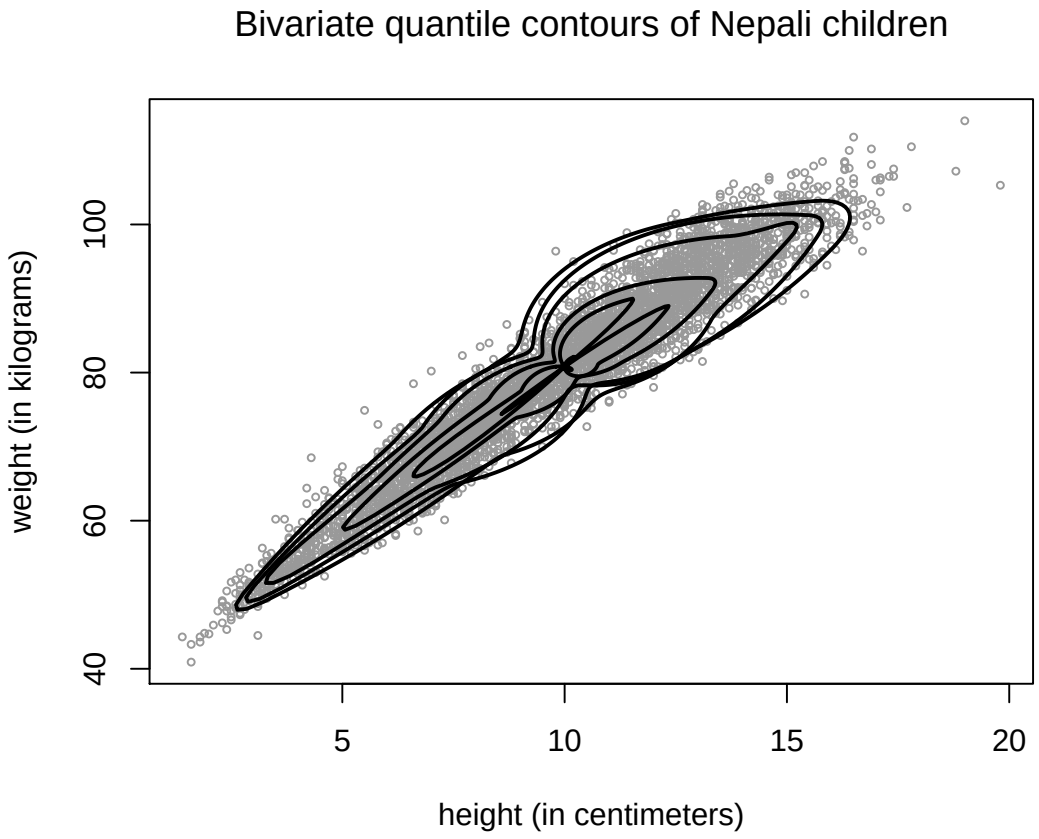

Figure 1.4: *The quantile contours, proposed by Wei (2008), of Nepali children weight and height with* $p = 1 - \{0.03, 0.05, 0.10, 0.25, 0.40\}$ *were shown. Even with carefully chosen* μ*, we may still obtain not so ideal contours. The contours for small* p*-values may be self-crossing.*

assessment of children's growth status. In Wei's method, there is still one issue which has not been fully addressed, the choice of the parameters μ and S. As recommended by Wei, we usually would choose μ to be the component-wise median and take the scale parameters in S as the component-wise median absolute deviation (MAD) for the purpose of robustness. The reason we stick to component-wise standardizations is that each component usually has its own implications. Although in general the derived

quantiles are not affine, nor orthogonally equivariant, they are equivariant with respect to translation and coordinate-wise rescaling (thanks to preliminary normalization).

As pointed out by Wei, the quantile contours should be applied with caution if the support of the random variable X is restricted to a certain subset of $\mathbb{R}^2$, for instance, banana-shaped data set, demonstrated in the appendix of Wei (2008). However, even if it is not, we should be very careful as well. The shape of quantile contours are significantly affected by the choice of the center μ, as shown in Figure 1.4. Even with carefully chosen μ, we may still obtain not so ideal contours. The self-crossing feature for small p-values is not appealing either. In the calculation, we use the constrained smoothing algorithm of He and Ng (1999) for convenience in estimating the quantile g, also the algorithm used in Wei (2008).

1.4 Problems considered in the thesis

Growth charts are one of most widely used tools in medical practice. They identify subjects that are unusual, in the sense that particular measurements, like height or weight, are extreme on the charts. A typical growth chart, as seen in Figure 1.5, consists of a family of smooth curves representing a few selected percentiles of the distribution of some physical measurements (height, weight, head circumferences, etc.) of the reference population as a function of age. In practice, univariate growth charts are usually fitted using one of two major approaches. One of them relies on the assumption that the response measurement is normally distributed, or can be transformed to normality: for example by Box-Cox transformation (Cole, 1988; Cole and Green, 1992). The other approach is more nonparametric; Wei, Pere, Koenker and He (2006) use nonparametric quantile regression to construct univariate growth charts.

Given the widespread usage of univariate growth charts, health practitioners (Pere,

Figure 1.5: *Weight-for-age growth charts of U.S. girls aged from* 2 *to* 20 *years; the percentiles are* $\{3, 10, 25, 50, 75, 90, 95\}\%$.

2006) started to look for bivariate growth charts, a natural generalization of univariate growth charts. These charts would be capable of visualizing two response measurements (for example, height and weight) depending on a covariate (age, say). Apparently, a task like this is quite complex and can be approached from various directions. One of the straightforward proposals based on the assumption of a bivariate normal distribution of two response variables was made by Thompson and Fatti (1997), (see

also Petersen, 2003). In fact Thompson and Fatti (1997) proposed bivariate reference region without covariates. The problem with this approach is that the joint distribution of two response measurements, for example height and weight, is rarely normal.While the transformation to normality may work reasonably well in the univariate case, in the bivariate case it is much harder to achieve desired flexibility and quality of fit. These reasons led Wei (2008) to propose a new form of multivariate quantile contours, avoiding strong distributional assumptions and taking time effect into account via a two-stage estimation procedure. Her multivariate quantile contours are flexible enough to include potentially important covariates and allow for probabilistic interpretation: the coverage of the centered p-th contour is p.

Wei's method, despite its undisputable merits, also raises some problems. The contours are significantly affected by the choice of the location parameter μ and scale parameter S. As Wei (2008) admits, the selection of these parameters is subject to debate. For small p values, the contours may be self crossing, even in the absence of covariates. Despite these shortcomings, the proposal of Wei (2008) is capable of delivering much of the required objective, and compares favorably to possible approaches based on extensions of certain existing multivariate quantile concepts.

In this thesis, we propose bivariate growth charts based on directional quantile envelopes. Directional quantile envelopes are the intersection of quantile regression lines in all directions, they are essentially depth contours when no covariates are involved. Compared to Wei (2008), their probabilistic interpretation has different character: while the contours of Wei (2008) interpret in terms of *coverage mass*, directional quantile contours admit an interpretation in terms of *tangent mass*. Therefore, even if our proposal lacks certain shortcomings of that of Wei (2008), both proposals are rather complementary. Finally, directional quantile envelopes share many

common virtues with the halfspace depth contours, as originally proposed by Tukey (1975) for plotting bivariate data, they enhance these by allowing for adding covariates, and hence provide a promising technique with diverse applications.

1.5 Overview of the results

The results of this thesis start from Chapter 3 following the review of various quantile and quantile regression concepts in Chapter 2. In Chapter 3 we introduce a concept of directional quantiles, which are proved to be continuous with respect to distributions in the Pompeiu-Hausdorff distance sense and with respect to directions. We provide two methods to visualize directional quantiles: quantile biplots and directional quantile envelopes. We concentrate on the latter: directional quantile envelopes, the intersection of the supporting halfspaces determined by directional quantiles. We derive their closed-form expression for elliptically symmetric distributions in Theorem 3.3.4. The main contribution of this chapter is the introduction of multivariate quantiles via a directional approach.

The main result of Chapter 4 is Theorem 4.3.1, " the retrieval theorem". This theorem provides explicit probabilistic interpretations of directional quantile envelopes: the probability mass of any tangent halfspace of the p-th directional quantile envelope is approximately equal to p; and when the tangent halfspace is unique the probability mass is exactly p. The theorem indicates that directional quantiles not only give a rank to multidimensional data but also provides meaningful interpretations. In Chapter 4, we also discuss the connection between directional quantile envelopes and halfspace depth contours (Tukey, 1975). It is proved in Theorem 4.2.1 that the p-th directional quantile envelope is equal to the upper level set of the p-th halfspace depth contour. In the end of Chapter 4 we bring some progress in the so-called characterization

problem for halfspace depth. "The characterization theorem", Theorem 4.4.1, is derived for distributions with smooth directional quantile envelopes, the case that was not previously covered in the literature.

Chapter 5 is devoted to the estimates and asymptotic properties of directional quantile envelopes and multivariate quantile regression via directional quantile envelopes. We describe our estimation scheme of directional quantile envelopes and prove that they are affine equivariant if the estimators of the corresponding directional quantiles are affine equivariant in Theorem 5.1.1. As an example, we use our results to estimate bivariate extreme quantiles. One of the main contributions of the thesis is that we construct bivariate growth charts: an important application of multivariate quantile regression. We also compare our directional approach to bivariate growth charts with the method proposed by Wei (2008). In the reminder of Chapter 5, we prove two main theorems. The first one, Theorem 5.4.1, is about the consistency of the estimates of the directional quantile envelopes: under some conditions of the estimates of directional quantiles, the estimates of the directional quantile envelopes are consistent in the Pompeiu-Hausdorff distance sense. The second one, Theorem 5.4.2, describes how accurate the approximation of the estimates of directional quantile envelopes can be.

While Chapters 3-5 develop the theoretical and practical aspects of multivariate quantile regression, we also discuss some further problems in the theory of halfspace depth. A part of this is their computation, in which we develop a fast elimination algorithm in Chapter 6. The algorithm constructs the set of active halfspaces whose intersection forms a directional quantile envelope. Applying this algorithm to a large number of quantile halfspaces, we can construct an arbitrary exact approximation of the direction quantile envelope.

Chapters 3-6 are devoted to the development and theory of multivariate quantile

regression applied to growth charts. The remaining two chapters contain additional results of the thesis. Chapter 7 is devoted to an alternative, more complicated than that we eventually adopted, for constructing multivariate quantile regression. Nevertheless, it is of some interest. In Theorem 7.2.1, we exhibit the connection between halfspace depth contours and directional regression quantiles (Laine, 2001), stated without proof in Koenker (2005). Our proof uses the duality theory of primal-dual linear programming. In Chapter 7, we also explore and prove some properties of the structure of halfspace depth contours for empirical distributions, absolute continuous distributions and certain general distributions, which help to explicitly interpret depth contours.

Chapter 8 is an execution to the general theory of quantiles in various statistical models. We propose a generalized quantile concept, depth quantile, aiming at generalizing some virtues of halfspace depth into quantile settings, inspired by halfspace depth (Tukey, 1975) and regression depth (Rousseeuw and Hubert, 1999). In the case of multivariate location, we prove that the quantiles are affine equivariance in Theorem 8.3.1 and give their half closed-form expression when the population distribution is elliptically symmetric in Theorem 8.3.2. We investigate the relationship to the special case of univariate location and present the result in Theorem 8.4.1. In regression with intercept, we clarify the relationship to the original Koenker's depth-based regression quantile in Theorem 8.5.2. We point out certain differences arising in regressions without intercept and gives an example showing that while the L_1-based quantile regression of Koenker and Bassett (1978) fails our quantile regression provides sensible answers.

A condensed version of Chapters 3-5, co-authored with Dr. Ivan Mizera, has been submitted for publication (arXiv:0805.0056v1, 2008). Chapter 6 with an `R` package, co-authored with Dr. Ivan Mizera, will be submitted for publication.

Chapter 2

Review of quantile concepts

This chapter contains an overview of selected statistical concepts involving quantiles. We review various multivariate quantile concepts from the literature, as well as give a short introduction to quantile regression. Central to the future development is the discussion, on page 27, of desirable properties of multivariate quantiles. We start our review from the univariate case.

2.1 Univariate quantiles

Recall Definition 1.1.1; for $0 < p < 1$ the p-th quantile of a distribution P is defined to be

$$Q(p) = \inf\{x : \ F(x) \geq p\},$$

where $F(x) = P((-\infty, x])$ is the cumulative distribution function of P. However, some authors may prefer to use F^{-1} instead of Q to denote the quantile function (see Csorgo, 1983 and Shorack, 2000). To make it less confusing, in the rest of the thesis we will only use the notation Q to denote the quantile function defined in Definition 1.1.1.

The *support* of the cumulative distribution function F of a random variable X with distribution P is defined to be the minimal closed set C such that $P(X \in C) = 1$; we can denote $suppX = C$. The concept of support can be easily extended to random

variables in high dimension. A *realizable p-th quantile* (Shorack 2000) of F, for $0 < p < 1$, is any value x such that $F(x) = p$. Cumulative distribution function F of any random variable X is right continuous; as its "inverse", quantile function Q is left continuous. Quantile function provides a complete characterization of a random variable X, in the same manner as cumulative distribution function F. Moreover we have the following two properties on F and Q.

Proposition 2.1.1 *(i) For any cumulative distribution function F and its corresponding quantile function Q, we have*

$$F(Q(p)) \geq p \quad \textit{for all } 0 < p < 1,$$

and equality fails if and only if $p \in (0, 1)$ is not in the range of F on its support, that is, a realizable p-th quantile does not exist.
(ii) For any cumulative distribution function F and quantile function Q, we have

$$Q(F(x)) \leq x \quad \textit{for all } -\infty < x < \infty,$$

and equality fails if and only if $F(y) = F(x)$ for some $y < x$, that is, the realizable quantile is not unique.

The proof of Proposition 2.1.1 is trivial and will not be provided here. For those who are interested, see Shorack (2000). Essentially, we tend to view Q as a function of p inverse to F, that is, a function satisfying

$$F(Q(p)) = p. \tag{2.1}$$

However, a simplified definition via the identity (2.1) would work only in regular cases, F is continuous; the sophistication of Definition 1.1.1 is needed to handle situations when there is none, or more than one $Q(p)$ satisfying (2.1). In this connection, it is

useful to invoke the following alternative quantile definition via minimization of the so called "check" function, $\rho_p(x) = x(p - I(x < 0))$ (see Ferguson, 1967).

Definition 2.1.1 *For $0 < p < 1$, the p-th* quantile set *of a distribution P is defined to be the set $\mathcal{Q}(p)$ of all q minimizing the integral*

$$\int \rho_p(x - u)P(dx),$$

where $\rho_p(x) = x(p - I(x < 0))$ is the "check" function. Every prescription that results in a unique element of every p-th quantile set—like that given by Definition 1.1.1—will be referred to as a quantile version.

Definition 2.1.1 exploits a well-known fact that every quantile set is a closed interval, possibly a singleton; the latter case takes place, in particular, when there is no $Q(p)$ satisfying (2.1), the case we refer to as *void* quantile. If, on the other hand, the set of q such that $F(q) = p$ is nonempty, then this set is equal to the quantile set; if it contains more than one element, we speak about *ambiguous* quantiles—"multiply realizable quantiles" in the terminology of Shorack (2000).

While working with a set-valued definition may have theoretical advantages, practitioners rather demand suitable quantile version: either the $\inf$, "theoretical" one given by Definition 1.1.1, or some other choice. Hyndman and Fan (1996) reviewed those in practical use, which are also implemented as options of the `R` function `quantile` (Frohne and Hyndman, 2004; `R` Development Core Team, 2008). For the computations in this thesis, we used the "`type=1`" version of `quantile`, abiding by the theoretical Definition 1.1.1, with the objective to produce maximally faithful illustrations of the explained theoretical facts. In concrete applications, we might rather consider one of the interpolated versions—say, the `R` default.

It is not hard to see that quantile ambiguity can occur only when the distribution contains a "gap", an open interval such that $P((a, b)) = 0$, and quantile voidness if the

distribution contains an atom, a point c such that $P\{c\} > 0$. These phenomena can be often ruled out for population distributions—and although inevitable for empirical ones, it should be noted that their extent often vanishes with growing sample size.

Let F_n be the empirical cumulative distribution of F. F_n *converges in distribution* to F if $F_n(x) \to F(x)$ at each continuity point x of F, denoted by $F_n \xrightarrow{d} F$. Let Q_n and Q denote the quantile function associated with cumulative distribution function F_n and F, respectively. We say that Q_n *converges in quantile* to Q if $Q_n(p)$ converges to $Q(p)$ at each continuity point p of Q in $(0, 1)$, denoted by $Q_n \xrightarrow{d} Q$.

It is known that any cumulative distribution function has at most countably infinite number of discontinuities. It is easy to show that any quantile function has at most countably infinite number of discontinuities as well, see Shorack (2000). The following proposition is not difficult and plays an important role in empirical quantile process, which reveals the equivalence in convergence relationship between quantile function and cumulative distribution function. The proof can be found in Shorack (p.112, 2000), and will not be provided here.

Proposition 2.1.2 *Convergence in distribution is equivalent to convergence in quantile, that is*

$$F_n \xrightarrow{d} F \text{ if and only if } Q_n \xrightarrow{d} Q.$$

2.2 Univariate quantile regression

To avoid measure-theoretic notation, we will only consider the sample data case. However, we speak about population quantiles instead of their estimators. Suppose that we have a group of observation data $(y_1, y_2, \cdots, y_n)$ of univariate response variable Y. Definition 2.1.1 of univariate quantile says that the *unconditional quantiles* $Q(p)$ of

Y, for $0 < p < 1$, is

$$Q(p) = \arg\min_{u} \sum_{i=1}^{n} \rho_p(y_i - u),$$

where $\rho_p(\cdot)$ is the "check" function as defined in Definition 2.1.1. Suppose that $(x_1, x_2, \cdots, x_n)$ are observed values of an explanatory variable X. Similarly, as in classical regression for the conditional mean, we formulate a functional relationship $g(x)$ for the conditional quantile of Y given X. The analog of the regression estimate is then given by the following definition.

Definition 2.2.1 *Given a functional model, the p-th regression quantile of Y conditionally on X is given as the solution of*

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_p(y_i - g(x_i)),$$

where ρ_p is the "check" function $\rho_p(x) = x(p - I(x < 0))$ (Koenker and Bassett, 1978).

A well known special case of regression quantile is the least absolute deviation (LAD) estimator when $p = 0.5$, corresponding to the conditional median fit. LAD is a robust alternative of ordinary least squares (OLS) estimator. In ordinary least squares regression model, regression curve only provides a grand summary for the mean of the response Y conditionally on covariates X. One may be interested, however, in getting more complex information from the data: for instance, the extreme behavior of Y conditionally on X. Quantile regression computes several different regression curves corresponding to various percentiles of the response and in this way obtains a more complete picture of the set, see Hogg (1975), and Koenker and Hallock (2001).

As we mentioned above, g can be fitted within a parametric model or nonparametric model. As an example to illustrate the methodology in the parametric setting, we fit a simple linear regression model, $g(x_i) = \beta_0 + \beta_1 x_i$. The data we use is the salaries of full and associate professors from the top 50 universities in the USA (the data set

is available at *http://lib.stat.cmu.edu/DASL/Datafiles/FacultySalaries.html*). In Figure 2.1, we can see that there is an obviously linear relationship. We superimposed five estimated quantile regression lines corresponding to $p = \{.05, .25, .50, .75, .95\}$. The least squares regression line is plotted as the dashed line. The plot clearly reveals the tendency of the dispersion of full professor salaries to increase along with its level as associate salaries increase. However, the spacing of the quantile regression lines also reveals that the conditional distribution of full professor salaries is skewed to the right:

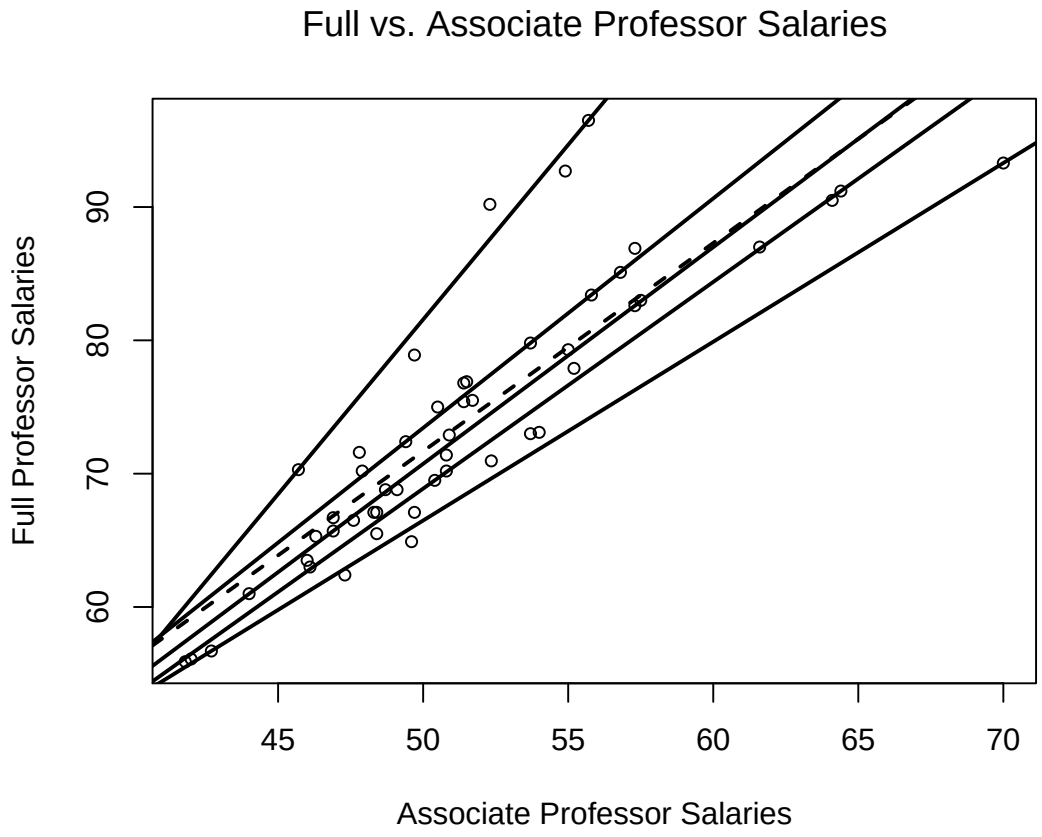

Figure 2.1: *The dashed line is the ordinary least squares mean regression; the solid lines are quantile regression corresponding to* $p = .95, .75, .25, .05$*. Quantile regression provides a more complete picture of the data set.*

the wider spacing of the quantiles indicating a low density and long lower tail and the narrower spacing of the quantiles indicating a higher density and a shorter upper tail.

Semi-parametric and nonparametric techniques of quantile regression have recently received a lot of attention; for overviews, see Buchinsky (1998) and Koenker (2005). Nonparametric fitting is especially useful when the functional relationship between response variable and explanatory variable is hard to specify explicitly; for example, the relationship between height or weight and age in univariate growth charts (Wei and He, 2006); see Figure 5.7 and Figure 5.6. Nonparametric methods for quantile regression include kernel-based approaches, regression and penalized splines, and others. Quantile regression (linear, nonlinear, and nonparametric) is implemented in the R package `quantreg` (Koenker, 2007) and the SAS PROC `QUANTREG` (Chen, 2005).

For nonparametric quantile regression in this thesis, we used regression splines: we fitted quantile function g as a linear combination of a few fixed basis functions; the B-spline basis is particularly convenient for this purpose (Wei, Pere, Koenker and He, 2005). Given a choice of knots for the B-splines, estimation of the quantile function is a straightforward application of parametric linear quantile regression with $z_{ij} = \phi_j(x_i)$ where ϕ_j is the j-th function in the B-spline basis. Given measurements y_i and their corresponding covariates x_i, $i = 1, 2, \cdots, n$, we first select a set of knots $\{\pi_j\}$, $j = 1, 2, \cdots, m$. Using the covariates x_i and knots π_j, we can derive the B-spline basis functions ϕ_j, $j = 1, 2, \cdots, df$, where df is the degrees of freedom of the B-splines. The p-th quantile regression can be then obtained by minimizing

$$\min \sum_{i=1}^{n} \rho_p \left(y_i - \sum_{j=1}^{df} \phi_j(x_i)\beta_j \right).$$

Optimization of the above equation can be performed efficiently by linear programming techniques, even for very large data sets (Koenker, 2005).

Obviously, other smoothing methods, e.g., kernel smoothing, may be used in lieu of penalty smoothness and regression splines. However, one advantage of splines is that in typical applications, it is usually sufficient to pre-select a set of knots using our understanding of the regression quantile functions, which is appealing in practice. For example, it would be useful to place more knots during infancy and puberty than during other times in the application of univariate growth charts, Wei and He (2006) and Wei, Pere, Koenker and He (2006).

2.3 Multivariate quantiles: an overview

Motivated by the success of univariate quantiles, researchers started to think about analogous extensions of univariate quantiles to other models. The most targeted situation in this respect was the multivariate quantiles. While multivariate quantiles may not be that interesting per se, their understanding is important as the first step to quantile regression with multivariate response. The lack of a natural basis for ordering multivariate data has always been a serious obstacle to the development of statistical methods based on order statistics. Therefore, the extension of univariate quantiles to multivariate quantiles poses a serious problem.

Usually, for any concept of multivariate quantiles, there is an associated multivariate median. In other words, we first have the concept of a multivariate median and then extend it to multivariate quantiles. There are many methods to define a multivariate median, see Small (1990) for a review; however, some of them are not capable of being extended into multivariate quantiles. For some multivariate medians, different authors gave various multivariate quantile extensions in distinct aspect with different interpretations, for example, spatial L_1 median, the Oja median and so on. However, not every multivariate extension of quantiles derives from a multivariate median; sometimes,

multivariate quantiles are generalized directly from univariate quantiles: approaches based on norm minimization. To provide some criteria and more perspectives for evaluating and comparing concepts of multivariate quantiles to be reviewed in the next sections, we list several basic questions for any version of them.

1. Do the multivariate quantiles support *probabilistic interpretations* analogous to that of the univariate quantiles?

2. Do the multivariate quantiles support formulations of multivariate notions of *nonparametric descriptive statistics for location*, such as median, trimmed means, and other L-statistics?

3. Do the multivariate quantiles satisfy any equivariance properties, such as *affine equivariance*, or *orthogonal equivariance*?

4. Do the multivariate quantiles possess any robust properties, such as bounded *influence function* or positive *breakdown point*?

5. Do the multivariate quantiles have any practical algorithm to be implemented to calculate them?

6. Do the multivariate quantiles have the capability to be extended to multivariate quantile regression models?

In what follows, we will review various quantile concepts paying special attention to how they confirm to the desideratum above. The classification of the review is partially adopted from Serfling (2002).

2.3.1 Approaches based on norm minimization

For a univariate random variable X with distribution P, if its expectation $E[X] = \int xP(dx) < \infty$, then the p-th quantile $Q(p)$ defined by Definition 2.1.1 for $0 < p < 1$

may be characterized as any value u minimizing

$$\int \phi(p, x-u)P(dx) \tag{2.2}$$

(See Ferguson, 1967, p. 51), where $\phi(p,u) = |u| + (2p-1)u$. Two generalizations of univariate quantiles to multivariate quantiles were started from (2.2).

The first generalization was derived by Abdous and Theodorescu (1992) by applying $\phi(p,u)$ on each element of a d-dimensional random variable X. Given a sample $\{x_1, x_2, \cdots, x_n\}$ with $x_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T, i = 1, 2, \cdots, n$, the p-th quantile Q(p), for $0 < p < 1$, is obtained by solving

$$\min_{q \in \mathbb{R}^d} \sum_{i=1}^{n} \left\| \frac{1}{2}\phi(p, x_{i1} - q_1), \cdots, \frac{1}{2}\phi(p, x_{id} - q_d) \right\|_r , \tag{2.3}$$

where $q = (q_1, q_2, \cdots, q_d)^T$, $r \in [1, \infty)$ is pre-selected, $\|\cdot\|_r$ denotes the usual L^r-norm on $\mathbb{R}^d$ and $\phi(p,u) = |u| + (2p-1)u$ defined above. For fixed r, the multivariate quantiles defined by (2.3) generate a series of *curves* in $\mathbb{R}^d$, with "central" and "extreme" quantiles corresponding, respectively to small and large values of $|p - 1/2|$. For $r = 1$, the multivariate quantiles are simply the vector of the marginal univariate p-th quantiles of all components, namely, coordinate-wise quantiles. In this case, the multivariate quantiles is only translation equivariant but not even orthogonal equivariant. For $r = 2$, the spatial median is obtained when $p = 0.5$. In general, $Q(p)$ is only orthogonal equivariant but not fully affine equivariant. Despite any possible mathematical appeals of (2.3) as a generalization of (2.2), what makes it less appealing is that this notion fails to conform to any satisfactory interpretation and hence can not provide meaningful rank for multivariate data.

The concept of multivariate quantiles given by Chaudhuri (1996) as a generalization of (2.2) is based on extending $\phi(p,u)$ in a different way. The idea is to index d-dimensional multivariate quantiles by elements of an open unit ball $B^d = \{p : p \in$

$\mathbb{R}^d, \|p\| < 1\}$, where $\|\cdot\|$ is the Euclidean distance. For any $p \in B^d$ and $u \in \mathbb{R}^d$, let $\phi(p,u) = \|u\| + p^T u$, then the proposed "geometric" quantile $Q(p)$ corresponding to the index $p \in B^d$ based on d-dimensional data $\{x_1, x_2, \cdots, x_n\}$ is defined as

$$Q(p) = \arg\min_{q \in \mathbb{R}^d} \sum_{i=1}^{n} \phi(p, x_i - q). \tag{2.4}$$

Geometric quantiles defined by (2.4) possess not only magnitude but also direction, as well as the multivariate quantiles proposed by Abdous and Theodorescu (1992). Serfling (2002) pointed out that they have a robustness property in the sense that $Q(p)$ remains unchanged if for fixed p any point x_i moves outward along $x_i - Q(p)$. However, geometric quantiles are only orthogonal equivariant but not affine equivariant. Due to this, they do not lead to any sensible estimate when different coordinate variables of the data-vectors are measured in different units or they have different degrees of statistical variations.

Chakraborty (2001) proposed a technique, transformation-retransformation method, to modify geometric quantiles defined by (2.4) to make them affine equivariant. First the geometric quantiles were extended to L_r quantiles. Let $\phi_r(p,u) = \|u\|_r + p^T u$ where $\|\cdot\|_r$ is L_r-norm with $r \geq 1$, the L_r quantile $Q^{(r)}(p)$ corresponding to $p \in B^d$ based on data $\{x_1, x_2, \cdots, x_n\}$ is defined as

$$Q^{(r)}(p) = \arg\min_{q \in \mathbb{R}^d} \sum_{i=1}^{n} \phi_r(p, x_i - q). \tag{2.5}$$

Given n data points $\{x_1, x_2, \cdots, x_n\}$ in $\mathbb{R}^d$, assuming that $n > d+1$, let $\alpha = \{i_0, i_1, \cdots, i_d\}$ denote a subset of size $(d+1)$ of the index set $\{1, 2, \cdots, n\}$. Consider the points $x_{i_0}, x_{i_1}, \cdots, x_{i_d}$, which form a data-driven coordinate system, where x_{i_0} determines the origin and the lines joining the origin to the remaining d data points $x_{i_1}, \cdots, x_{i_d}$ form various coordinate axes. The $d \times d$ matrix $X(\alpha)$ containing the columns $x_{i_1} - x_{i_0}, \cdots, x_{i_d} - x_{i_0}$ can be taken as the transformation matrix for

transforming the remaining data points x_j's, $1 \leq j \leq n$ and $j \notin \alpha$ to $z_j^{(\alpha)} = [X(\alpha)]^{-1}x_j$, expressing the data points in terms of the new coordinate system. Under some regular conditions, the transformation matrix $X(\alpha)$ will be an invertible matrix with probability one. Let

$$\upsilon(\alpha) = \begin{cases} \dfrac{[X(\alpha)]^{-1}p}{\|[X(\alpha)]^{-1}p\|_q}\|p\|_q, & \text{for} \quad p \neq 0 \\ 0, & \text{for} \quad p = 0, \end{cases}$$

where q satisfies $1/r + 1/q = 1$. Then the transformation-retransformation (TR) L_r quantile $Q^{(\alpha,r)}(p)$ is defined as the solution of

$$\min_{q \in \mathbb{R}^d} \sum_{i \notin \alpha} \left\{ \left\| [X(\alpha)]^{-1}(x_i - q) \right\|_r + \{\upsilon(\alpha)\}^t \{X(\alpha)\}^{-1}(x_i - q) \right\}. \tag{2.6}$$

Chakraborty (2001) proved that the TR L_r quantiles defined by (2.6) are equivariant under arbitrary affine transformations of the data. For any vector $z \in \mathbb{R}^d$, we are able to define invariant notations of multivariate ranks based on the transformation-retransformation approach. For example, in the approach if we use L_2-norm

$$n^{-1} \sum_{x_i \neq z, i \notin \alpha} \left\| [X(\alpha)]^{-1}(x_i - z) \right\|_2^{-1} [X(\alpha)]^{-1}(x_i - z)$$

can be view as descriptive statistics that determine the geometric position of the point $z \in \mathbb{R}^d$ in terms of the observed data cloud. Similarly, for any other L_r-norms, we can construct the corresponding versions of multivariate ranks based on different transformation matrix $X(\alpha)$. With the derived multivariate ranks for each observed data point, multivariate Q-Q plot can be obtained to compare the closeness of two independent samples (Chakraborty, 2001). However, the performance of the quantile estimators depends on the choice of the transformation matrix $X(\alpha)$. Different choices of the indexing subset may result in different transformation matrices, hence maybe distinct rankings. It is important to select a suitable subset of indices α. The adaptive approach proposed by Chakraborty and Chaudhuri (1998) was claimed to be capable of

picking up a subset with good properties. Nevertheless, the dependence on the selection of subset indices α raises a question about numerical stability of the transformation-retransformation approach, is still an issue needing more investigation.

2.3.2 Spatial quantiles

In the univariate case, the p-th quantile can be rewritten as the scalar that minimizes the sum of residuals weighted by an appropriate loss function, ρ_p; see Definition 2.1.1. This idea can be extended to define univariate M-quantiles by using a weighted form of a primitive of the Huber's M-function (Huber, 1964) as the loss function. The difficulty in a multivariate adoption of this concept is mainly caused by the fact that any natural ordering in $d > 1$ dimension does not exist. As proposed by Breckling and Chambers (1988), we need to introduce a directional unit vector s first, that is, $s \in \mathbb{S}^{d-1} = \{s : \ \|s\| = 1, \text{ and } s \in \mathbb{R}^d\}$, which will be used as a direction index. Given the datapoints $\{x_1, x_2, \cdots, x_n\}$ where $x_i \in \mathbb{R}^d$ and quantile index $0 < p < 1$, let α_i denote the angle between $x_i - q$, $q \in \mathbb{R}^d$, and $s \in \mathbb{S}^{d-1}$. The p-th M-quantile (Breckling and Chambers, 1988) $Q^s(p)$ is the vector solving

$$\min_{q \in \mathbb{R}^d} \sum_{i=1}^{n} \left(1 - (1-2p)\cos\alpha_i\right) \rho(x_i - q),$$

where

$$\rho(u) = \begin{cases} \dfrac{1}{2c}\|u\|^2 & \text{if} \quad \|u\| < c, \\ \|u\| - \dfrac{c}{2} & \text{if} \quad \|u\| \geq c; \end{cases} \tag{2.7}$$

and $\|\cdot\|$ is the Euclidean distance. The basic idea of this extension is to introduce a weighting scheme for the residual $x_i - q$ according to not only its magnitude and also the angle between it and s. In the univariate case, the definition reduces to the ordinary M-quantile where either $s = 1$ or $s = -1$. The parameter $c \geq 0$ determines where this weighting changes from an expectile type weighting to an ordinary quantile

type weighting. When s moves around the whole unit sphere $\mathbb{S}^{d-1}$, the resulting set of corresponding M-quantiles is a $(d-1)$ dimensional closed surface embedded within the d dimensional Euclidean space $\mathbb{R}^d$.

While the above definition yields acceptable results in many situations, the result can be in certain circumstances very different from what appears to be intuitive. A reasonable natural quality of a multivariate quantile is that it lies within the convex hull of the sample data in the light that multivariate quantiles are capable of providing ranks of the data. This, however, is not always the case as the following example shows. We consider a two-dimensional "cigar shaped" data set; see Figure 2.2. The sample was generated by adding normally distributed random terms with mean 0 and variance 0.1 to 200 equidistant data points on the interval $[-1, 1]$. To keep the results simple, we set $c = 0$ in (2.7) and compute the multivariate quantiles with $p = \{.1, .2, .3, .5\}$ and s moving around the whole unit circle. The quantiles, which construct a contour if they possess the same p value, are showed in the left panel in Figure 2.2, where we can see most of the quantiles are lying outside the convex hull constructed by the sample data points even for a reasonably large p.

Considering that the multivariate M-quantiles defined by (2.7) are based on some loss functions, Breckling, Kokic and Lubke (2001) chose an alternative way to define M-quantiles by directly using appropriate influence functions to make the quantiles lying within the data convex hull. Under the above notations, the revised M-quantiles $Q^s(p)$ (Breckling, Kokic and Lubke, 2001) can be written as the solution of

$$\sum_{i=1}^{n}(x_i - q)w_i = 0$$

where

$$w_i = \begin{cases} \dfrac{1-(1-2p)\cos\alpha_i}{c} & \text{if} \quad \|x_i - q\| < c \\ \dfrac{1-(1-2p)\cos\alpha_i}{\|x_i - q\|} & \text{if} \quad \|x_i - q\| \geq c \end{cases} \tag{2.8}$$

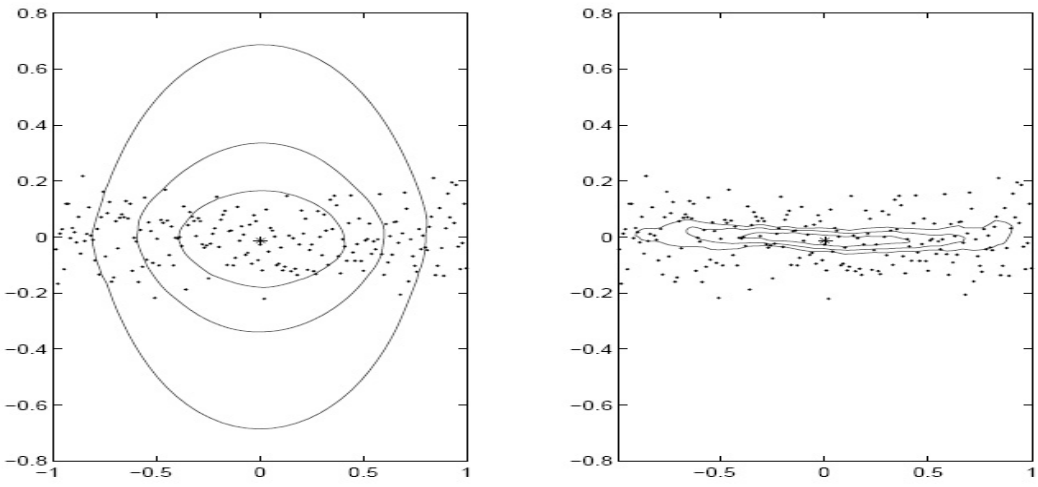

Figure 2.2: *The left panel is the M-quantiles ($p = \{.1, .2, .3, .5\}$) proposed by Breckling and Chambers (1988), most of which lie outside the data convex hull. The right panel is the M-quantiles ($p = \{.1, .2, .3, .5\}$) revised by Breckling, Kokic and Lubke (2001) and all of them lie within the data convex hull.*

As expected, the quantiles defined by (2.8) reduce to the standard M-quantiles in the univariate case. It is easy to check that setting c to be zero yields the spatial median and setting it to be infinity yields the multivariate sample mean. The revised M-quantiles perform more intuitively than the original ones proposed by Breckling and Chambers (1988). Comparatively, in the right panel of Figure 2.2 we illustrate the revised M-quantiles using the same data set as in the left panel with the same settings. We can see all the quantiles lie within the convex hull constructed by the sample data points. An M-quantile as defined by (2.7) and (2.8) is equivariant under rotation and translation transformations, but not equivariant under affine transformation.

Koltchinskii (1997) studied the M-quantiles and proposed a different definition in another aspect. Given a sample of data points $\{x_1, x_2, \cdots, x_n\} \in \mathbb{R}^d$ and a strictly convex function $f(\cdot, x) : \mathbb{R}^d \to \mathbb{R}$, the Young-Fenchel conjugate of $f_n(u) =$

$\sum_{i=1}^{n} f(u, x_i)$ is defined as

$$f_n^\star(p) = \sup_{u \in \mathbb{R}^d} \left[u^T p - f_n(u) \right],$$

where $p \in \mathbb{R}^d$ is quantile index. For suitable p, the p-th M-quantile proposed by Koltchinskii (1997) is defined as a subgradient of $f_n^\star(p)$. For instance, in the univariate case where $d = 1$, we can let $f(u, x) = |u - x| - |x|$, then we will obtain the usual univariate quantiles; see Definition 2.1.1. Koltchinskii (1997) pointed out that the subgradient of $f_n^\star(p)$ may not be unique; however, under some mild conditions of $f(u, x)$ and the sample data, the uniqueness of the subgradient can be guaranteed. To simplifying the discussion, we will assume that all the subgradients of $f_n^\star(p)$ are unique in what follows.

Given a strictly convex function $f(\cdot, x)$, let $F_n = \partial f_n$, then we have $F_n = \sum_{i=1}^{n} \partial f(u, X_i)$ and $\partial f_n^\star = F_n^{-1}$. F_n and F_n^{-1} are called M-distribution function and M-quantile function, respectively. In multivariate cases, we can define spatial quantiles using the kernel function

$$f(u, x) := \varphi(|u - x|) - \varphi(|u_0 - x|),$$

where $u, x \in \mathbb{R}^d$; $u_0 \in \mathbb{R}^d$ is a fixed point and φ is a convex nondecreasing function on $[0, \infty)$ and differentiable in $(0, \infty)$. The M-quantiles proposed by Koltchinskii (1997) have many good mathematical properties, for example, they are equivariant under translation and orthogonal transformation, not fully affine equivariant though. Meanwhile they can be used to extend L-parameters and L-estimators to the multivariate cases. By choosing proper kernel function $F(u, x)$, the corresponding "median" $F^{-1}(0)$ is the well-known Haldane's spatial L_1 median. For more properties and applications, see Koltchinskii (1994,1996, and 1997).

2.3.3 Quantiles based on the Oja median

Let $\{x_1, x_2, \cdots, x_n\}$ be a sample of data points coming from $\mathbb{R}^d$, the volume of the d-dimensional simplex determined by $x_{i_1}, x_{i_2}, \cdots, x_{i_{d+1}}$ is defined as,

$$\triangle(x_{i_1}, x_{i_2}, \cdots, x_{i_{d+1}}) = \left|\det[(\mathbf{1}, x_{i_1}, x_{i_2}, \cdots, x_{i_{d+1}})^T]\right| / (d+1)!$$

where $\mathbf{1} = (1, 1, \cdots, 1)^T \in \mathbb{R}^d$ and $\det[A]$ denotes the determinant of the matrix A. The Oja median (Oja, 1983) μ_o^α is obtained by solving the following equation,

$$\min_{\mu \in \mathbb{R}^d} \sum \left[\triangle(x_{i_1}, \cdots, x_{i_d}, \mu)\right]^\alpha, \quad 0 < \alpha < \infty, \tag{2.9}$$

where the summation runs over every indices satisfying $1 \leq i_1 < i_2 < \cdots < i_d \leq n$. For example, for $d = 2$ we sum over every pairs of data points.

Deke (2000) extended the Oja median to Oja multivariate quantiles, in which the sign of the determinant plays a key role in determining quantiles, while in (2.9) it was eliminated by taking absolute value. In the proposal of Deke (2000), we first need to rearrange the data points $x_{i_1}, x_{i_2}, \cdots, x_{i_d}$ such that their first elements are listed in ascending order. When first elements of data points are equal, we consider the second elements until the arrangement is completed. Unless there are two identical data points where the volume of the simplex is zero, this arranging is unique. Without loss of generality, we still denote them by $x_{i_1}, x_{i_2}, \cdots, x_{i_d}$ while we should keep in mind that they have been sorted already. For $p \in (0, 1)$ the p-th quantile $Q(p)$ (Deke, 2000) based on the Oja median is defined as

$$Q(p) = \arg\min_{q \in \mathbb{R}^d} \sum \rho_p \left(sign(\det[(\mathbf{1}, x_{i_1}, x_{i_2}, \cdots, x_{i_{d+1}})^T]) \triangle(x_{i_1}, \cdots, x_{i_d}, q)\right) \tag{2.10}$$

where ρ_p is the "check" function defined in Definition 2.1.1, that is $\rho_p(x) = x(p - I(x < 0))$; see Figure 2.3 for quantiles defined by (2.10) with $p = \{.75, .85, .95\}$.

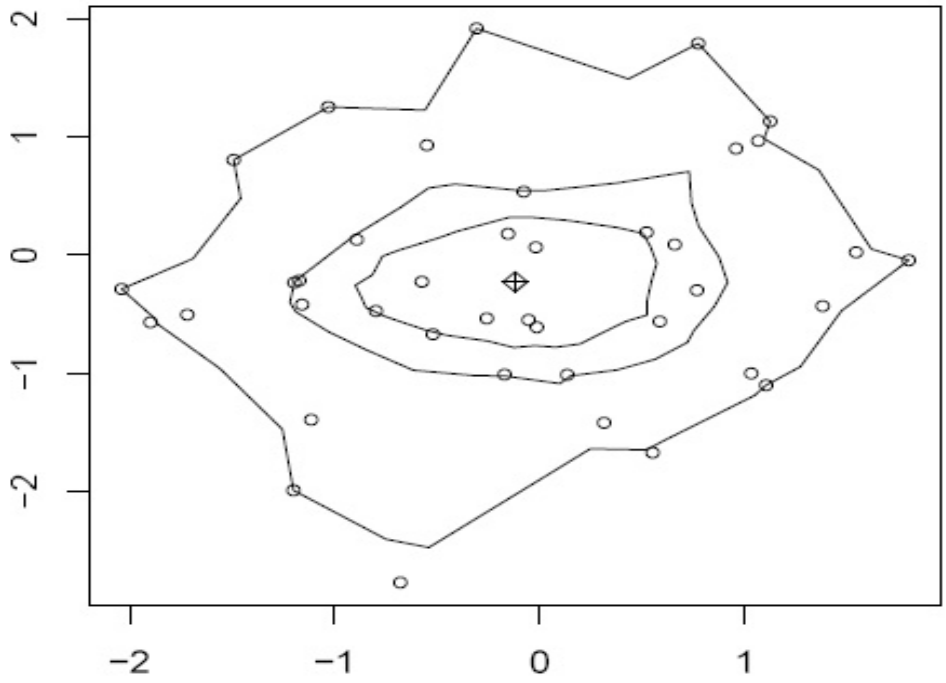

Figure 2.3: *The center point marked by diamond with crossing symbol is the Oja median. The contours are the Oja quantile contours with different p values ($p = \{.75, .85, .95\}$). Clearly, the contours are not convex.*

One of the advantages of these Oja quantiles is that they are affine equivariant. However, computation of these quantiles proved to be very difficult due to the curse of dimensionality. Ronkainen, Oja and Orponen (2001) proposed some methods to compute the Oja multivariate median, which could be modified to compute the Oja multivariate quantiles (Koenker, 2005). While the influence function is bounded, the Oja median has 0% breakdown point which means it is not robust. A fundamental difficulty with the approach is that there is no apparent way to assess the probability content of the resulting contours. Clearly, the approach presumes a unimodal nominal

model and this too may be questioned in many circumstances. In addition, from Figure 2.3, we can see that the quantile contours are highly affected by the shape of data and are not necessarily convex.

2.3.4 Generalized quantile processes

An approach to capture particular scalar features of a multivariate distribution by means of designed univariate type "generalized quantile" functions is formulated by Einmahl and Mason(1992). Let $\mathcal{A}$ be a subset of Borel set $\mathcal{B}$ in $\mathbb{R}^d$ and λ be a real-valued function defined on $\mathcal{A}$, the generalized quantile function based on an empirical probability measure P_n, subset $\mathcal{A}$ and quantile function λ is defined as follows:

$$Q(p) = \inf_{A \in \mathcal{A}} \{\lambda(A) : P_n(A) \geq p\}, \quad p \in (0,1). \tag{2.11}$$

In univariate case, let $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ and λ be the Lebesgue measure on the real line(*i.e.,* , $\lambda((-\infty, x]) = x$), $Q(p)$ defined by (2.11) is the classical real-valued one-dimensional quantile function defined by Definition 1.1.1. The definition by (2.11), with the possibility of various choices of $\mathcal{A}$ and λ, makes generalized quantiles a very attractive tool when dealing with multivariate data (Serfling, 2002). The asymptotic performance of the corresponding empirical quantile process was also studied by Einmahl and Mason(1992).

Liu, Parelius and Singh (1999) proposed to use quantile functions to rank multivariate data, which essentially are a type of general quantiles (Serfling, 2002). Given a nonnegative real-valued depth function $d(x, P_n)$ where $x \in \mathbb{R}^d$ and P_n is an empirical distribution function, Liu, Parelius and Singh (1999) suggested plotting the volumes of the corresponding p-th central regions,

$$V(p, d, P_n) = \texttt{volume of } C(p, d, P_n), \quad p \in (0,1), \tag{2.12}$$

versus p to construct a simple one-dimensional "scale curve". Multivariate distributions may be studied individually or comparatively in terms of these scale curves. Serfling (2002) pointed out that $V(p, d, P_n)$ is a generalized quantile function. Specifically, let $\mathcal{A}$ be the class of inner regions of $d(\cdot, P_n)$, denoted by $I(\alpha, d, P_n)$, such that $I(\alpha, d, P_n) = \{x : d(x, P_n) \geq \alpha\}, \alpha \geq 0$ and $\lambda(I(\alpha, d, P_n))$ be volume of $I(\alpha, d, P_n)$, we obtain $V(\alpha, d, P_n)$, $p \in (0, 1)$, which is $Q(p)$ defined by (2.11). For more discussion and applications on generalized quantile processes, see Serfling (2002).

In this section, we just attempt to review and compare several different multivariate quantile approaches conceptually, which represent distinct typical directions. There are many other multivariate quantile generalizations, which can not be covered due to space limitations. For other possible generalizations of quantile, see Small (1990) and Serfling (2002).

2.4 Multivariate quantile regression

To detect the dependence on the covariates in the lower and upper tails of the response variables and obtain a complete figure of the dependence, regression quantiles are very useful tools in regression models with univariate response variables. However, the extension to regression models with multivariate response variables, that is, multivariate quantile regression, still poses a problem. Only a few multivariate quantiles reviewed in Section 2.3 are capable of being extended in regression model.

Chakraborty (2003) proposed a concept of multivariate quantile regression based on the transformation and retransformation L_r quantiles defined by (2.6). Suppose that we have a sample of n data points $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$ coming from (X, Y), where $X \in \mathbb{R}^k$ and $Y \in \mathbb{R}^d$; we assume that $n > d + k$. Let S_n be the collection of all subsets of $k + d$ indices from the set $\{1, 2, \cdots, n\}$. For a fixed

index set $\alpha = \{i_1, \cdots, i_k, j_1, \cdots, j_d\} \in S_n$, let $W(\alpha)$ be the $k \times k$ matrix whose columns are the vectors $x_{i_1}, \cdots, x_{i_k}$ and $Z(\alpha)$ be the $d \times k$ matrix whose columns are the vectors $y_{i_1}, \cdots, y_{i_d}$. We assume that $W(\alpha)$ is invertible, which is guarantied with probability one under some regular conditions, and define $E(\alpha)$ to be the $d \times d$ matrix consisting of the columns $y_{j_1} - Z(\alpha)[W(\alpha)]^{-1}x_{j_1}, \cdots, y_{j_d} - Z(\alpha)[W(\alpha)]^{-1}x_{j_d}$. Under some mild conditions, $E(\alpha)$ is also invertible with probability one. For any $p \in B^d = \{p : p \in \mathbb{R}^d, \|p\| < 1\}$, where $\|\cdot\|$ is the Euclidian distance or L_2-norm, let

$$V(\alpha) = \begin{cases} \dfrac{[E(\alpha)]^{-1}p}{\|[E(\alpha)]^{-1}p\|}\|p\|, & \text{for} \quad p \neq 0 \\ 0, & \text{for} \quad p = 0, \end{cases}$$

then given an index set α, the p-th TR (Transformation-Retranformation) quantile regression can be obtained by solving

$$\min_{\beta \in \mathbb{R}^{d \times k}} \sum_{i \notin \alpha} \left\{ \left\| [E(\alpha)]^{-1}(y_i - x_i^T\beta_p) \right\| + V(\alpha)^T[E(\alpha)]^{-1}(y_i - x_i^T\beta_p) \right\} \tag{2.13}$$

where $\|\cdot\|$ is the L_2-norm. If $k = 1$ and $x_i = 1$ for $i = 1, 2, \cdots, n$, then the regression quantiles defined by (2.13) reduce to the TR L_2 multivariate location quantiles introduced by Chakraborty (2001) as in (2.6). Under general conditions, the TR quantile regression estimators exist and are uniquely determined (Chakraborty, 2003). The TR regression quantiles are not only regression equivariant but also affine equivariant for response variables, inheriting from their unconditional counterparts, multivariate location quantiles. However, a similar issue of multivariate regression quantiles to multivariate location quantiles will arise: how to select the index subset α to make the estimation stable.

Another concept of multivariate regression quantiles was introduced by Breckling and Chambers (1988). Recall that there is always an influence function ψ_p associated with the loss function ρ_p for any M-quantiles. Given a sample $\{y_1, y_2, \cdots, y_n\}$ of response

variables $Y \in \mathbb{R}^d$ and covariates $\{x_1, x_2, \cdots x_n\}$ of $X \in \mathbb{R}^k$, the quantile regression based on M-quantiles (Breckling and Chambers, 1988) can be obtained by solving

$$\sum \psi_p(y_i - x_i^T \beta_p) = 0, \quad \sum \psi_p(y_i - x_i^T \beta_p)^T x_i = 0. \tag{2.14}$$

Breckling and Chambers (1988) provided two examples on the application of the quantile regression in 2.14. The multivariate quantile regression based on M-quantile inherits some good properties from M-quantiles. However, there are several important issues which need to be addressed, for example, how to choose the influence function ψ_p, how does ψ_p effect the regression, and most important, how to interpret the regression results. There is an analogue multivariate quantile regression based on the multivariate M-quantiles proposed by Koltchinskii (1997). That quantile regression may possess some good mathematical properties, but its interpretation and applications need exploration. We will not go into detail here, see Koltchinskii (1997) for more information.

While these and other multivariate quantile regression proposals—for example, Deke (2000) extended the quantiles based on the Oja median to multivariate quantile regression—may at times enjoy certain favorable properties, like robustness, rotation equivariance, even affine equivariance etc., the real question is that when going to applications, whether they satisfy our needs. The answer here is not that simple, but we believe that much of it will be clarified in the next chapters.

2.5 Halfspace depth and depth contours

As we mentioned earlier, due to the lack of natural ordering in multidimensional space, there is no universally preferred definition of multivariate quantiles. Among the various ideas on multivariate quantiles developed in the literature in order to rank data, the crucial point is to find a way to rank the data points in high dimension. In this sense,

depth is able to provide a center-outward ordering of data points and makes a paramount important contribution to rank data in multidimensional space, where halfspace depth, or Tukey depth, may be one of the the most successful representatives. The commonly used name " Tukey depth" reflects that Tukey (1975) proposed depth contours for plotting bivariate data, even though Hodges (1955) first introduced it.

Tukey (1975) introduced the following notion of the centrality of a point x with respect to a distribution P.

Definition 2.5.1 *The* halfspace depth *of a point $x \in \mathbb{R}^d$ with respect to a probability distribution P is*

$$d(x, P) = \min_{\|u\| \neq 0} P\{y : \ u^T y \geq u^T x, \ y \in \mathbb{R}^d\}.$$

In the above expression, if we restrict $\|u\| = 1$, $u^T x$ becomes the projection of x onto the direction u. In this sense, $d(x, P)$ can be rewritten as

$$d(x, P) = \min_{\|u\| \neq 0} P\{y : \ (y - x) \in H_u, \ y \in \mathbb{R}^d\}, \tag{2.15}$$

where $H_u = \{x : \ u^T x \geq 0\}$ is the closed halfspace that contains 0 on its boundary, with vector u pointing inside the halfspace and orthogonal to the boundary. The point(s) with maximum halfspace depth provide a measure of centrality known as the Tukey median; it is an alternative to population mean, but more resistant to outliers; it can be considered as a generalization of the population median considering that it is population median in one dimension. The breakdown point of the Tukey median is at most $1/(d + 1)$, where d is the dimension, and can be as high as $1/3$ for a centrally symmetric distribution; see Mizera and Müller (2004). The sample halfspace depth of a point x with respect to an empirical distribution P_n based on data $\{y_1, y_2, \cdots, y_n\}$ is

$$d(x, P_n) = \inf_{\|u\| \neq 0} \#\{i : \ u^T y_i \geq u^T x, \ y_i \in \mathbb{R}^d\}/n = \inf_{\|u\| \neq 0} \#\{i : \ (y_i - x) \in H_u\}/n,$$

where P_n is the empirical probability distribution, $\#\{\cdot\}$ denotes the number of data points in $\{\cdot\}$ and H_u is defined as in (2.15). In what follows we will use $d(x)$ as an abbreviation of $d(x, P)$ or $d(x, P_n)$ if no confusion arises.

As the concept of halfspace depth is capable of providing a rank of data points in multidimensional space, we would expect it satisfying some equivariant properties, such as *translation equivariance* and *affine equivariance*, that is, halfspace depth is independent of the coordinate system. By Definition 2.5.1, we can easily obtain that halfspace depth is translation and affine equivariance in the following sense.

Proposition 2.5.1 *Suppose that P_1 and P_2 are two probability distributions in $\mathbb{R}^d$. If for any $\mathcal{A} \subset \mathbb{R}^d$, $P_1(\mathcal{A}) = P_2(B\mathcal{A} + b)$, where $B\mathcal{A} + b = \{Bx + b : \ x \in \mathcal{A}\}$, B is a $d \times d$ nonsingular matrix and b is a vector in $\mathbb{R}^d$, then*

$$d(Bx + b, P_2) = d(x, P_1),$$

for any $x \in \mathbb{R}^d$.

In the case of empirical distributions, Proposition 2.5.1 still holds. Given two empirical distributions P_n and G_n in $\mathbb{R}^d$, if $P_n(\mathcal{A}) = G_n(B\mathcal{A} + b)$ for any $\mathcal{A} \subset \mathbb{R}^d$, then

$$d(Ax + b, G_n) = d(x, P_n).$$

For $p \in (0, 1)$, the p-th depth contour D_p is the collection of $x \in \mathbb{R}^d$ such that $d(x) \geq p$, that is,

$$D_p = \{x : \ d(x) \geq p \text{ and } x \in \mathbb{R}^d\}, \tag{2.16}$$

although a stricter usage might reserve this phase for the boundary of D_p (He and Wang, 1997). The deepest nonempty contour is the collection of the Tukey median. Depth contours are the analog of quantile contours, just as depth is the analog of quantiles. Immediately, we have the following properties about depth contours; see

Donoho and Gasko (1992). For more properties about halfspace depth, also see Massé and Theodorescu (1994), and Mizera and Volauf (2002).

Proposition 2.5.2 *Suppose P is a probability distribution in $\mathbb{R}^d$ and D_p is defined as in* (2.16)*.Then the depth contours form a sequence of nested convex sets: each D_p is convex and $D_{p_1} \subset D_{p_2}$, if $p_1 \geq p_2$.*

Since halfspace depth was first introduced by Hodges (1955) and reinvestigated by Tukey (1975), it becomes a very popular topic. Nowadays halfspace depth plays an important role in multivariate data analysis and is the most popular one among various depth functions. Halfspace is a special case of tangent depth, introduced by Mizera (2002). Given a data set $z_n = \{(x_i, y_i), i = 1, 2, \cdots, n\} \in \mathbb{R}^d$ and a collection of criterial functions $F_1, F_2, \cdots, F_n$ defined on a space Θ parameterizing the fits, then the *tangent depth* (Mizera, 2002) of ϑ is defined as

$$d_T(\vartheta) = \inf_{u \neq 0} \#\{i : \nabla F_i(\vartheta) \in H_u\}/n = \inf_{u \neq 0} \#\{i : u^T \nabla F_i(\vartheta) \geq 0\}/n,$$

where $\nabla F_i(\vartheta)$ denotes the gradient of $F_i(\vartheta)$ and $H_u = \{x : u^T x \geq 0\}$ denotes the closed halfspace that contains 0 on its boundary, with the vector u pointing inside the halfspace and orthogonal to the boundary. The depth-based analog of median is customarily called maximum depth estimator in the literature; see Donoho and Gasko (1992). Different criterial functions end up with different depth functions. For example, let $F_i(\vartheta) = \frac{1}{2}\|z_i - \vartheta\|^2$ where $\|\cdot\|$ is the Euclidean distance, then the gradient of F_i is $\nabla F_i(\vartheta) = z_i - \vartheta$ and the corresponding tangent depth of ϑ can be rewritten as

$$d_T(\vartheta) = \inf_{u \neq 0} \#\{i : (z_i - \vartheta) \in H_u\}/n = \inf_{u \neq 0} \#\{i : u^T(z_i - \vartheta) \geq 0\}/n,$$

which is nothing but the halfspace depth of ϑ. If we pick the criterial function to be $F_i(\vartheta) = \frac{1}{2}(y_i - x_i^T\vartheta)^2$, then we reach the regression depth defined by Rosseeuw and

Hubert (1999). Mizera and Müller (2005) introduced the location-scale depth, where halfspace depth is also one of its special cases.

In view of the vast literature on multivariate analysis, we cannot adequately describe all the important developments on data depth in this thesis. The most closely related methods have been mentioned in this section, in the light that depth-based methodology may be viewed in part as a multivariate generalization of standard univariate quantile methods. For general discussion covering many examples of depth functions in current use, see Liu, Parelius and Singh (1999) and Serfling (2002).

Chapter 3

Multivariate quantiles via a directional approach [1]

3.1 Directional quantiles

To exemplify how quantiles may be useful with multivariate data, let us look at a bivariate example, the weight and height of 4291 Nepali children, aged between 3 and 60 months—the data constituting a part of the Nepal Nutrition Intervention Project-Sarlahi (NNIPS, principal investigator Keith P. West, Jr., funded by the Agency of International Development). Obviously, using weight or height alone is not enough to provide complete information about the growth status of children, see Figure 3.1. One possible way, illustrated in Figure 3.1, is to invoke Quetelet's *body mass index* (hereafter *BMI*), defined as the ratio of weight to squared height (in metric units; imperial sources may include an adjusting multiplier). A nutrition expert may dispute the relevance of *BMI* for young children, and remind us of possible alternatives—for instance, the Rohrer index (ratio of weight to *cubed* height, hereafter *ROI*), as in Figure 3.1. However, we do not think that the problem lies in deciding whether that or another index is preferred; we may simply look at the quantiles of some suitable combination of weight and height. Quantiles of certain functions of variables (in

[1] A condensed version of Chapters 3-5, co-authored with Dr. Ivan Mizera, has been submitted for publication (arXiv:0805.0056v1, 2008).

particular, linear combinations) may provide valuable information about multivariate data, and that plotting the corresponding (directional) quantile lines is an appealing way to present this information—in particular, to indicate how the quantiles divide the data.

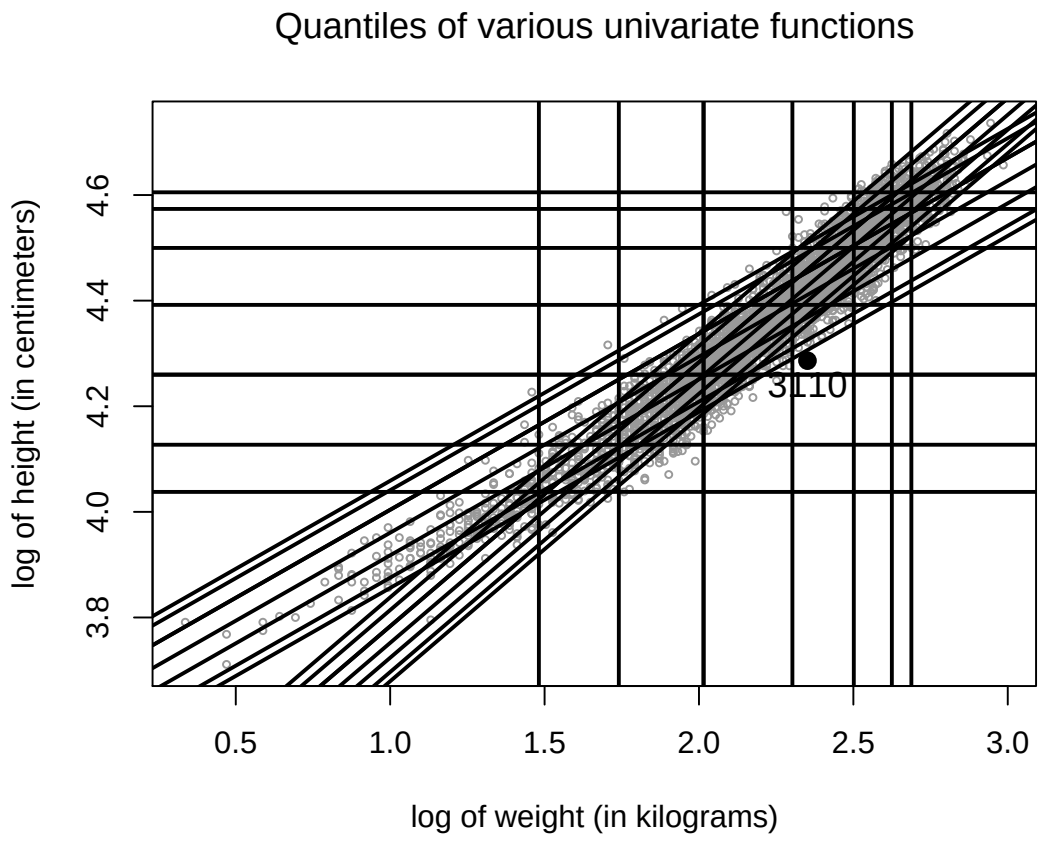

Figure 3.1: *Multivariate data typically offers insight beyond the marginal view, often through the quantiles of univariate functions of primary variables. Plotting the corresponding quantile lines is an appealing way to present this information.*

Notationally, it is often convenient to work with random variables/vectors, and write

$$Q(p) = Q(p, X) = \inf\{u : \mathbb{P}[X \leq u] \geq p\}, \tag{3.1}$$

despite the fact that quantiles depend only on the distribution, P, of a random variable

X. (The apparent notational convention hereafter is to suppress the dependence on X when no confusion may arise.) Equation (3.1) is another form of Definition 1.1.1.

Once we decide to focus on linear combinations, we realize that it is sufficient to look exclusively on projections; any other linear combination is a multiple of a projection, and the quantile of a multiple is the multiple of the quantile, which can be seen from Theorem 3.1.1. The following Definition 3.1.1 and Theorem 3.1.1 are elementary, but in a sense fundamental.

Definition 3.1.1 *An operator that assigns a point, or a set of points, T, to a random variable X is called* translation equivariant*, if its value for $X + b$ coincides with $T + b$, and* scale equivariant*, if its value for cX coincides with cT, where $c > 0$. (If T is a set, then the transformations are performed elementwisely.)*

Theorem 3.1.1 *Given a random variable X, for every $p \in (0, 1)$, the quantile operator $Q(p, X)$ is translation and scale equivariant.*

Proof. The translation equivariance follows from the identities

$$\begin{aligned} Q(p, X + b) &= \inf\{v : \mathbb{P}[X + b \le v] \ge p\} = \inf\{u + b : \mathbb{P}[X + b \le u + b] \ge p\} \\ &= \inf\{u : \mathbb{P}[X + b \le u + b] \ge p\} + b = \inf\{u : \mathbb{P}[X \le u] \ge p\} + b \\ &= Q(p, X) + b; \end{aligned}$$

where $v = u + b$ and $b \in \mathbb{R}$ is a constant.

Similarly, the scale equivariance follows from

$$\begin{aligned} Q(p, cX) &= \inf\{v : \mathbb{P}[cX \le v] \ge p\} = \inf\{cu : \mathbb{P}[cX \le cu] \ge p\} \\ &= c \inf\{u : \mathbb{P}[cX \le cu] \ge p\} = c \inf\{u : \mathbb{P}[X \le u] \ge p\} \\ &= cQ(p, X), \end{aligned}$$

where $v = cu$ and $c \in \mathbb{R}$ is an arbitrary positive constant.

Equivariance of any other quantile versions has to be checked individually—which is usually straightforward and will not be provided here. □

Informally, directional quantiles are the univariate quantiles of the projected distribution onto a normalized direction s—a vector with unit Euclidean norm; and the directional quantile lines are the lines with direction s and distance to the origin the same as the associated directional quantiles. Directional quantile lines indeed indicate how the corresponding directional quantiles divide the data. Mathematically, we have the following definition.

Definition 3.1.2 *We call any vector with unit Euclidean norm in $\mathbb{R}^d$ a* normalized direction*, and denote the set of all such vectors by $\mathbb{S}^{d-1}$, that is $\mathbb{S}^{d-1} = \{s : \ \|s\| = 1, \text{ and } s \in \mathbb{R}^d\}$. Let X be a random vector with distribution P. Given a normalized direction $s \in \mathbb{S}^{d-1}$ and $0 < p < 1$, the p-th* directional quantile*, in the direction s, is defined as the p-th quantile of the corresponding projection of the distribution of X,*

$$Q(p, s) = Q(p, s, X) = Q(p, s^T X) = \inf\{u : \ \mathbb{P}[s^T X \leq u] \geq p\}.$$

The corresponding p-th directional quantile hyperplane *(*directional quantile line *for $d = 2$) is given by the equation $s^T x = Q(p, s)$. The p-th* directional quantile set *is defined analogously as*

$$\mathcal{Q}(p, s) = \mathcal{Q}(p, s, X) = \mathcal{Q}(p, s^T X) = \arg\min \int \rho_p(s^T x - u) P(d(s^T x)),$$

where ρ_p is the "check" function defined in Definition 2.1.1.

The p-th directional quantile in the direction s and the $(1-p)$-th directional quantile in the direction $-s$ are not necessarily equal, due to the $\inf$ convention employed in Definition 1.1.1. Nonetheless, they often coincide—for instance, it is not possible to distinguish between any p-th and $(1-p)$-th directional quantile lines if quantile ambiguity, in the sense of Definition 2.1.1, is excluded for any projection of P. A sufficient condition for unambiguity is the following *contiguous support* property.

Definition 3.1.3 *Let P be a probability distribution in $\mathbb{R}^d$, of a random variable X. We say that P has* contiguous support*, if there is no intersection of any two halfspaces with parallel boundaries that has nonempty interior but zero probability and divides the support of P into two parts.*

Note that if the support is not contiguous, it will not be connected as well; however, it may be disconnected and still contiguous, for instance, if there is a stripe with nonparallel edges crossing the support. In this sense, the contiguous constraint is weaker than the disconnected constraint; the former contiguous support constraint was used by Mizera and Volauf (2002) and the latter disconnected support was used by Massé and Theodorescu (1994) where topics about continuity of halfspace depth were studied. When $d = 1$, in one dimensional space, two constraints are identical. We believe that the contiguous support, and thus the lack of quantile ambiguity, is a fairly typical virtue of population distributions, and consequently we will limit our attention to p from $(0, 1/2]$ (although some subsequent theorems will be formulated for more general p).

3.2 Continuity of directional quantiles

Continuity of directional quantiles plays an important role, not only in plotting, but also in theoretical study; hence is a basic requiring property and deserves some introspection. Recall the definition of directional quantile of $Q(p, s, X)$ in Definition 3.1.2, we can see it depends not only on quantile index p and random variable X, as usual, but also on direction s. In this section we will leave the continuities in p and X and focus our intention on the continuity of $Q(p, s, X)$ in terms of s.

Definition 3.2.1 *We say that a distribution P, of a random variable X in $\mathbb{R}^d$, has continuous p-th directional quantiles, if $Q(p, s, X)$ is a continuous function of $s \in \mathbb{S}^{d-1}$, for $p \in (0, 1)$.*

It turns out that this continuity is quite common; in particular, it takes place for all empirical distributions, which is a direct consequence of the following more general theorem. It also holds true for any other directional quantile version under certain conditions. Theorem 3.2.1 shows that the continuity of directional quantiles and the continuity of directional quantile sets, the continuities of $Q(p, s, X)$ and $\mathcal{Q}(p, s, X)$ in terms of s, is true for most random variables. In Theorem 3.2.1, when we say $\mathcal{Q}(p, s, X)$ is a *continuous* function of s, we mean $\mathcal{Q}(p, s_n, X)$ converges to $\mathcal{Q}(p, s, X)$ in the Pompeiu-Hausdorff distance as $s_n \to s$. Recall the definition of the Pompeiu-Hausdorff distance—the terminology we follow is that of Rockafellar and Wets (1998):

$$\begin{aligned} d(A, B) &= \max\big\{\sup_{x\in A} d(x, B), \sup_{y\in B} d(A, y)\big\} \\ &= \max\big\{\sup_{x\in A}\inf_{y\in B} d(x, y), \sup_{y\in B}\inf_{x\in A} d(x, y)\big\}, \end{aligned}$$

where A and B are two compact non-empty subsets of $\mathbb{R}^d$. Informally, the Pompeiu-Hausdorff distance between two sets of points, is the longest distance an adversary can force you to travel by choosing a point in one of the two sets, from where you then must travel to the other set.

Theorem 3.2.1 *Suppose that X is a random vector in $\mathbb{R}^d$ with distribution P. If the support of P is bounded, then $\mathcal{Q}(p, s, X)$ is a continuous function of s, for every $p \in (0, 1)$; in particular, $Q(p, s, X)$ is continuous in terms of s.*

The same results hold true when the support of X is contiguous; moreover, if a sequence of random vectors X_n converges almost surely to X, and $s_n \to s$, then $\mathcal{Q}(p, s_n, X_n)$ converges to $\mathcal{Q}(p, s, X)$ in the Pompeiu-Hausdorff distance, for every $p \in (0, 1)$; and $Q(p, s_n, X_n)$ tends to $Q(p, s, X)$ as well.

Proof. Since quantile sets are bounded intervals, it is sufficient to prove the convergence of their endpoints to $\inf \mathcal{Q}(p, s, X) = \inf\{u : \mathbb{P}[s^T X \le u] \ge p\}$ and $\sup \mathcal{Q}(p, s, X) = \sup\{u : \mathbb{P}[s^T X \ge u] \le (1 - p)\}$, respectively.

Suppose that the support of X is bounded. Let $q = Q(p, s, X) = \inf \mathcal{Q}(p, s, X)$, we have that $\mathbb{P}[s^T X \leq q] \geq p$ and $\mathbb{P}[s^T X \leq q - \epsilon] < p$ as a direct consequence of Definition 1.1.1, where $\epsilon > 0$. As the support of the random vector X is bounded, we have $\|X\| \leq M$ almost surely for some $M > 0$; by the Cauchy-Schwarz inequality, $|(s - s_n)^T X| \leq M\|s - s_n\|$ and therefore we have

$$p \leq \mathbb{P}[s^T X \leq q] = \mathbb{P}[s_n^T X \leq q - (s - s_n)^T X] \leq \mathbb{P}[s_n^T X \leq q + M\|s - s_n\|], \quad (3.2)$$

which means that $\inf \mathcal{Q}(p, s, X) \leq q + M\|s - s_n\|$. In a similar fashion, we obtain that $\inf \mathcal{Q}(p, s, X) \geq q - M\|s - s_n\| - \epsilon$, due to

$$\mathbb{P}[s_n^T X \leq q - M\|s - s_n\| - \epsilon] \leq \mathbb{P}[s^T X \leq q - \epsilon] < p. \quad (3.3)$$

Putting (3.2) and (3.3) together and letting $\epsilon \to 0$, we obtain

$$q - M\|s - s_n\| \leq \inf \mathcal{Q}(p, s_n^T X) \leq q + M\|s - s_n\|,$$

and therefore $\inf \mathcal{Q}(p, s, X) \to \inf \mathcal{Q}(p, s, X)$, that is, $Q(p, s_n, X_n)$ to $Q(p, s, X)$. The convergence of $\sup \mathcal{Q}(p, s, X) \to \sup \mathcal{Q}(p, s, X)$ is proved analogously.

If the support of a random vector X is contiguous, then all directional quantile sets in the limit are singletons. Pompeiu-Hausdorff convergence then follows from the "outer convergence" of quantile sets in the sense of Rockafellar and Wets (1998), see also Mizera and Volauf (2002): any limit point, x, of any sequence $x_n \in \mathcal{Q}(p, s_n, X_n)$ lies in $\mathcal{Q}(p, s, X)$. This can be easily seen in an elementary way, observing that $x_n \in \mathcal{Q}(p, s_n, X_n)$ entails

$$p \leq \limsup_{n \to \infty} \mathbb{P}[s_n^T X_n \leq x_n] \leq \mathbb{P}[s^T X \leq x]$$

and

$$1 - p \leq \limsup_{n \to \infty} \mathbb{P}[s_n^T X_n \geq x_n] \leq \mathbb{P}[s^T X \geq x]$$

Since under the contiguous support assumption the quantiles are unique, this second part of the theorem holds true for every quantile version. □

The theorem is formulated more generally, to allow for alternative quantile versions, and to facilitate later asymptotic considerations. Using the theorem with P to be an empirical distribution, hence possessing bounded support, shows that the continuity of directional quantiles holds for all empirical, and also many population distributions.

3.3 Directional quantile envelopes

Figure 3.2 shows the plot of the logarithms of weight and height of Nepali children, together with superimposed lines indicating quantiles, $p = \{.10, .20, .30, .40, .50, .60, .70, .80, .90\}$, in 20 uniformly spaced directions. While we still champion plotting directional quantile lines as an appealing way to present the directional quantile information, we have to admit that the plot becomes quickly overloaded if multiple directions and indexing probabilities are requested. Therefore, we would like to consider alternatives aimed at compression of the directional quantile information. While our focus is not exclusively graphical, the task of plotting is probably the most palpable one to epitomize this objective.

3.3.1 Quantile biplots

It looks as though the mere task of plotting of directional quantiles—for fixed p—should not pose any special difficulty: the directional quantile $Q(p, s)$ may be represented by the point $Q(p, s)s$ lying on the line with direction s and passing through the origin. Such a plot is more informative when superimposed on the original bivariate scatterplot; this leads to an amalgam which we decided to call a *quantile biplot*, and whose instances can be seen in Figure 3.3.

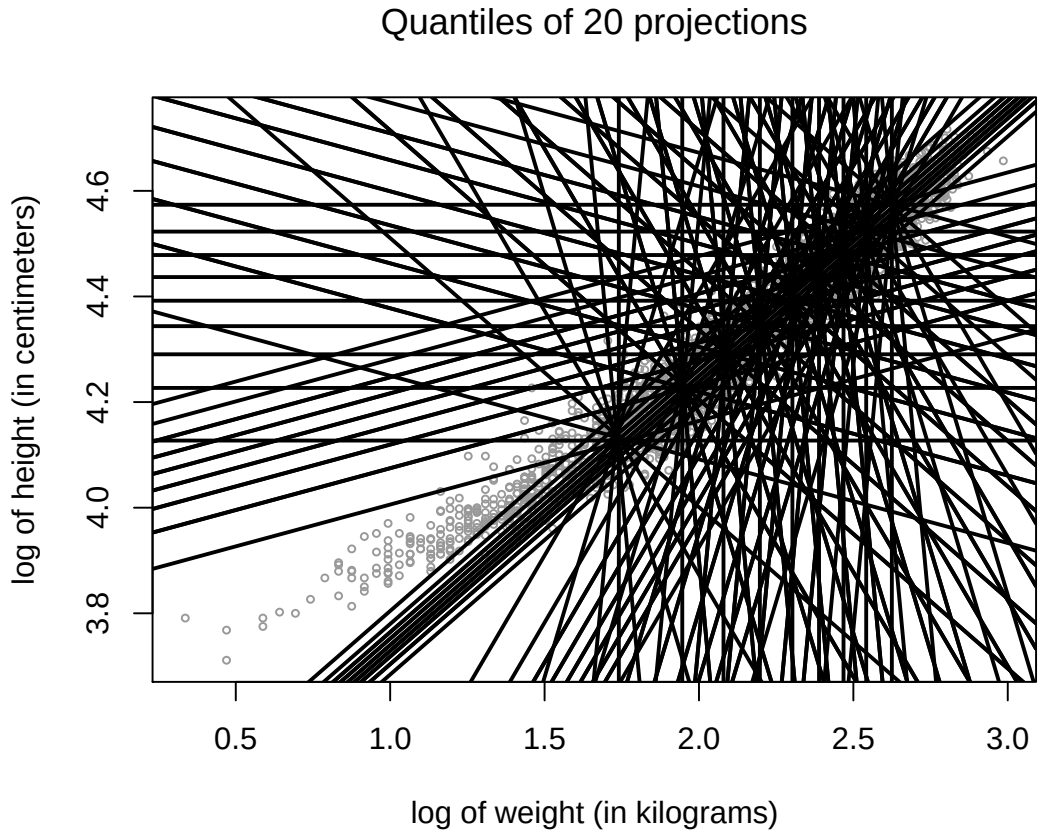

Figure 3.2: *The plot gets quickly overloaded if multiple directions and indexing probabilities are requested. Therefore, we would like to consider alternatives aimed at compression of the directional quantile information.*

Quantile biplots allow for faithful and relatively straightforward retrieval of the directional quantile information. Whenever a directional quantile line is sought, it is sufficient to find the intersection of the p-th contour with the halfline emanating from the selected origin in the given direction s starting in the selected origin. The line passing through the intersection and perpendicular to this halfline is then the desired directional quantile line, which divides the whole space into two halfspaces with probability mass about p and $1 - p$, respectively. Figure 3.3 shows how this works for the coordinate

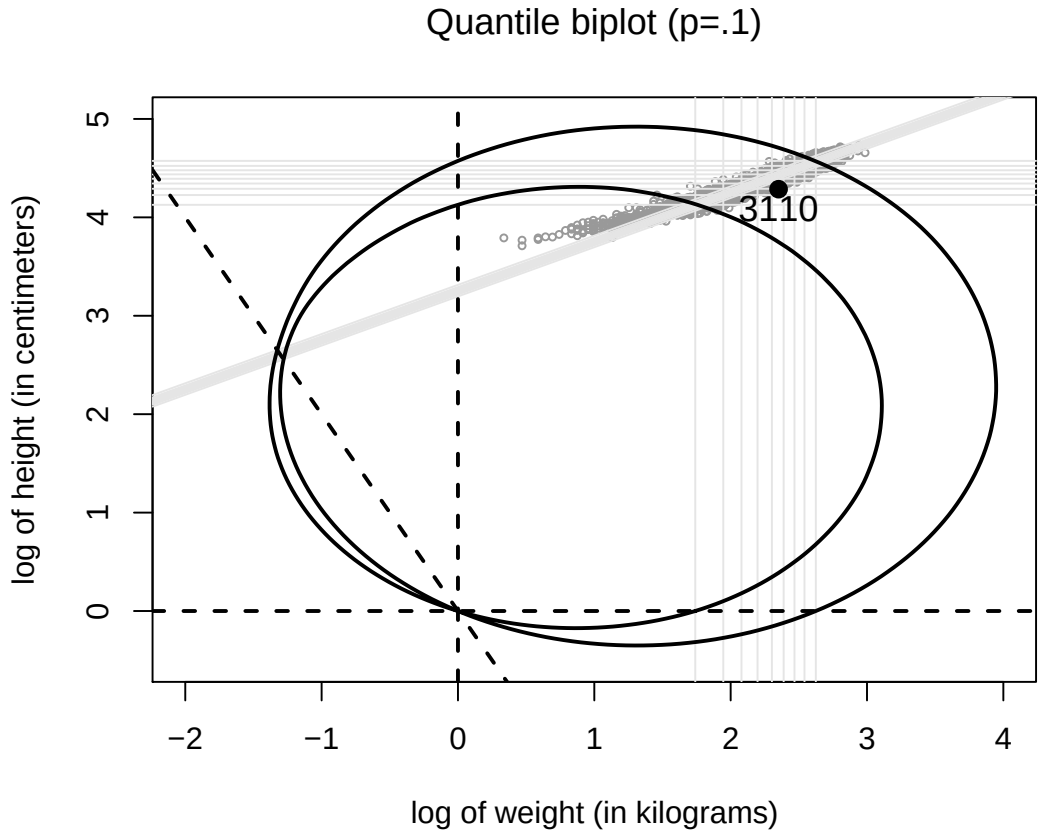

Figure 3.3: *Quantile biplots allow for faithful and relatively straightforward retrieval of the directional quantile information.*

directions, and also for the " *BMI* direction" $s \propto (-1/2, 1)$, where the directions are represented by dashed lines and the quantile lines are represented by grey lines with $p = \{.10, .20, .30, .40, .50, .60, .70, .80, .90\}$.

In Figure 3.3, we used the same coordinate system for both data and quantile ($p = .1$) components, the system inherited from the data. The origin thus happens to be located outside the data cloud; as a consequence, the "quantile contour" extends far away from the data cloud, and intersects itself. These characters of contours make quantile biplots

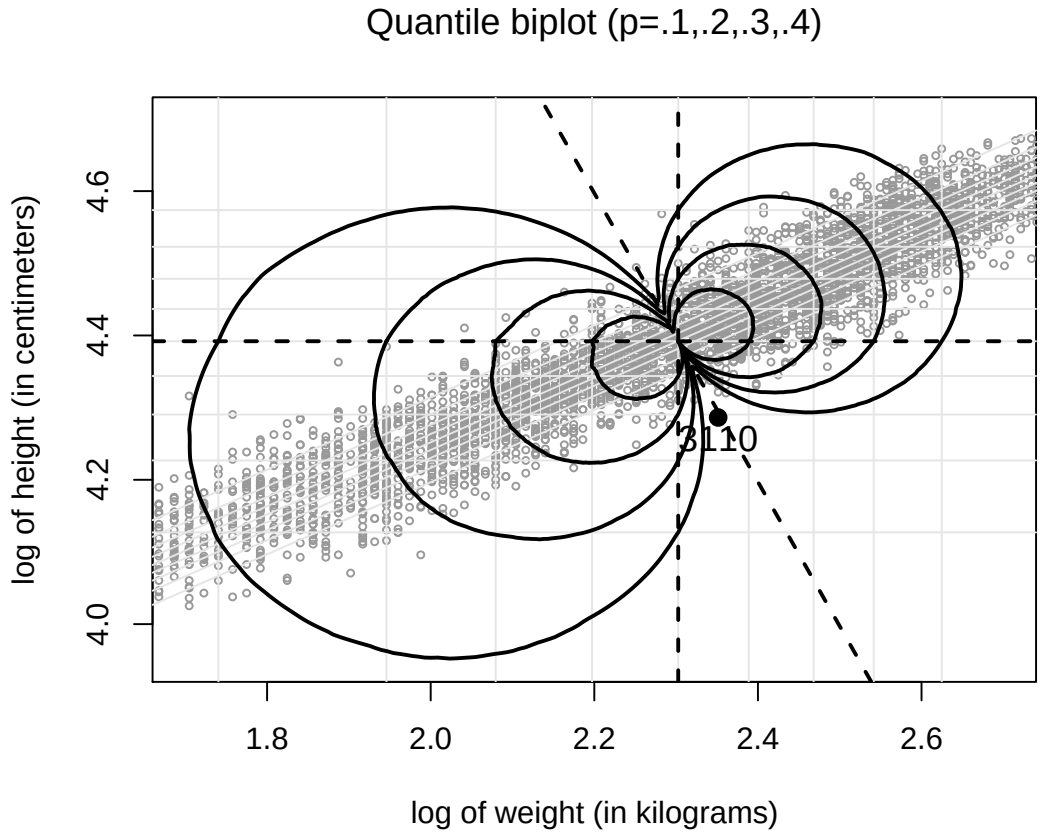

Figure 3.4: *The counterintuitive character of biplot contours, their dependence on the coordinate system, and certain other features (tendency to self-intersection and "mozzarella" shape) are rather disturbing.*

counterintuitive, and maybe less appealing. However, we can avoid some of these effects by choosing the origin for the quantile plotting inside the data cloud, the possibility shown in Figure 3.4, where we used the coordinate-wise median as the origin. To make it compatible with Figure 3.3, we transformed the plot back into the original coordinate system. Finding the appropriate location takes trial and error; the coordinate-wise median is a decent guess according to the following guideline for selecting the origin for plotting quantiles in quantile biplots—which may be perhaps of some interest, should

anybody find this way of plotting appealing. It turns out that such an origin should be as deep as possible; the best available one is thus located in the Tukey median, guaranteeing non-intersecting curves of quantile biplots for all p not exceeding its depth. In practice, the coordinate-wise median is a reasonable choice, as we did in Figure 3.4.

Theorem 3.3.1 *The curve of quantile biplots formed by $Q(p,s)s$ for fixed p and revolving s does not intersect itself if the origin of the coordinate system has halfspace depth no less than p.*

Proof. To show that the curve does not intersect itself, it is sufficient to prove that $Q(p,s) \leq 0$ for any $s \in \mathbb{S}^{d-1}$. As $p \leq d(0)$, $d(0)$ standing for the halfspace depth of the origin, for any $\epsilon > 0$ we have

$$P(\{x : s^T x \leq Q(p,s) - \epsilon\}) < p \leq d(0) \leq P(\{x : s^T x \leq 0\}),$$

where the first half part is from Definition 1.1.1 and the last half part is from the definition of halfspace depth. Therefore we have $Q(p,s) - \epsilon < 0$. Let $\epsilon \to 0$, we reach $Q(p,s) \leq 0$ for any $s \in \mathbb{S}^{d-1}$ and the proof is done. □

However, quantile biplots exhibit several disturbing features. One of them, as revealed in the search for the origin of a quantile biplot, is the lack of any equivariance—even with respect to an operation as simple as a shift. For plotting, the equivariance with respect to translations and coordinatewise rescaling is something like a minimum requirement—otherwise the automatic rescaling implemented in typical graphical routines may easily distort the plotted content.

Overall, quantile biplot contours appear rather counterintuitive, and their tendency to self-intersections and "mozzarella" shapes, as in Figure 3.4, probably will not being favorite. It seems that the question is not how to plot directional quantiles, but how to

successfully incorporate this information into the plot of the data. Directional quantile lines then appear still better than anything else—the only problem is to reduce the overload caused by their straightforward plotting.

3.3.2 Directional quantile envelopes

To convey the information carried by directional quantile lines while avoiding the plotting overload, we propose to take, for fixed $p \in (0, 1/2]$, the inner envelope of the p-th directional quantile lines. See Figure 3.5 for $p = .10$, where we take 100 evenly spaced directions and the directional quantiles are plotted in grey.

Definition 3.3.1 *Let X be a random vector in $\mathbb{R}^d$ with distribution P. For fixed $p \in (0, 1/2]$, the p-th* directional quantile envelope *generated by $Q(p, s)$ is defined as the intersection,*

$$D(p) = \bigcap_{s \in \mathbb{S}^{d-1}} H(s, Q(p, s)),$$

where $H(s, q) = \{x : \ s^T x \geq q\}$ is the supporting halfspace *determined by $s \in \mathbb{S}^{d-1}$ and $q \in \mathbb{R}^d$. The notation $D_A(p)$ will be used in case when the intersection is taken only over a subset $A \subseteq \mathbb{S}^{d-1}$ of all possible directions; in the spirit of this notation, $D(p) = D_{\mathbb{S}^{d-1}}(p)$.*

The nested feature of directional quantile envelopes is a direct consequence of Definition 3.3.1 considering that directional quantiles are monotone in any direction. In general, directional quantile envelopes are always bounded (if we are using a proper subset of directions to define $D_A(p)$; we usually want to ensure this by taking the direction set A not contained in any closed halfspace whose boundary contains the origin) and convex, being the intersection of convex sets. Therefore, we have the following theorem about directional quantile envelopes.

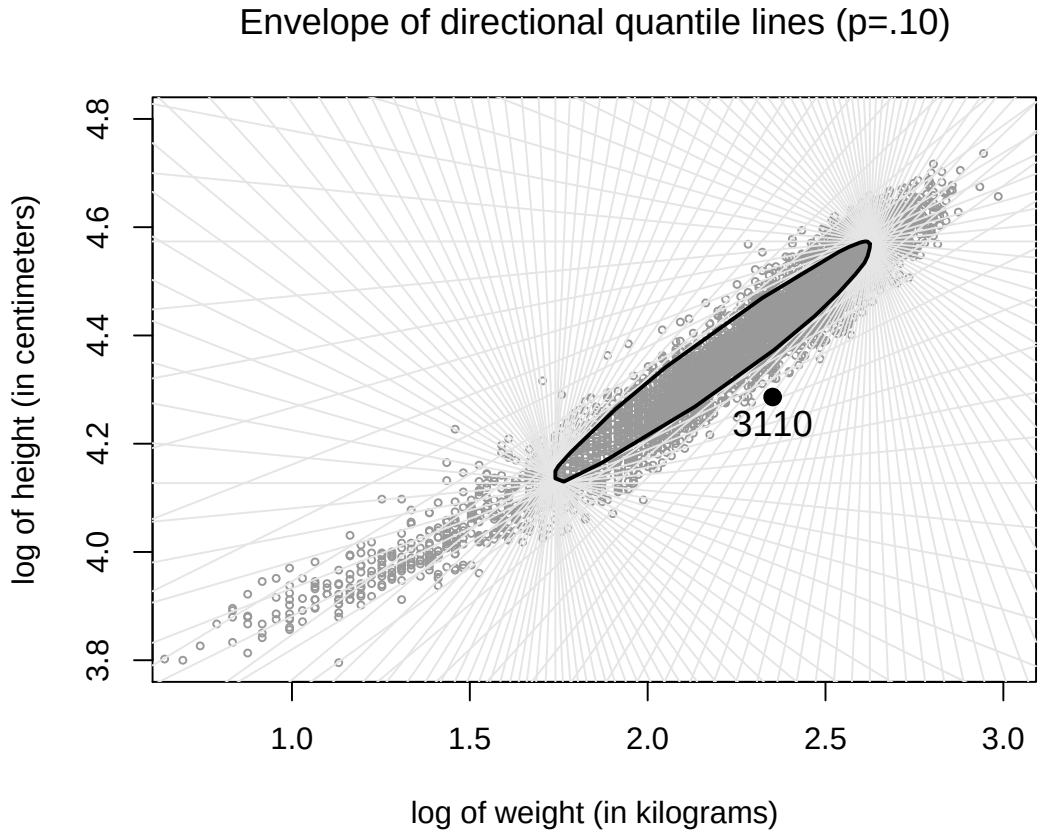

Figure 3.5: *We propose to take, for fixed $p \in (0, 1/2]$, the inner envelope of the p-th directional quantile lines.*

Theorem 3.3.2 *The directional quantile envelopes of the distribution P of a random variable X in $\mathbb{R}^d$ form a sequence of nested convex sets: each $D(p)$ is convex and $D(p_1) \subset D(p_2)$, if $p_1 \geq p_2$.*

The proof is very simple and will not be provided in the thesis. These features can also be observed in Figure 3.6. Another property worth pointing out is, if p is not so large, less than the maximum halfspace depth, $D_A(p)$ will be nonempty. See Theorem 4.2.1. Contrary to quantile biplots, directional quantile envelopes are affine equivariance. To

avoid any misunderstanding, let us refresh the definition of affine equivariance.

Definition 3.3.2 *An operator that assigns a point or a set T in $\mathbb{R}^d$ to a collection of data points $x_i \in \mathbb{R}^d$, is called* affine equivariant*, if its value is $BT + b$ when evaluated on the data points $Bx_i + b$, for any nonsingular matrix B and any $b \in \mathbb{R}^d$. (If T is a set, then the transformations are performed elementwise.)*

Theorem 3.3.3 *Given a random variable X in $\mathbb{R}^d$, directional quantiles of any version are translation and scale equivariant; and the directional quantile envelopes generated by these quantiles are affine equivariant.*

Proof. From Theorem 3.1.1, directional quantiles $Q(p, s, X)$ are translation and scale equivariant, for all $s \in \mathbb{S}^{d-1}$ and fixed p. These results hold true for any other directional quantile version. To simplifying the proof, we only consider directional quantiles $Q(p, s, X)$. However, the proof is the same for any other quantile version. Theorem 3.3.3 holds true for any quantile version.

Let B be a nonsingular matrix and b a vector. First, we verify that the transformation rule for the supporting halfspace of the directional quantile: for every $s \in \mathbb{S}^{d-1}$ and every $p \in (0, 1)$,

$$H(B^*s/\|B^*s\|, Q(p, s, BX + b)) = BH(s, Q(p, s, X)) + b, \tag{3.4}$$

where $B^* = (B^{-1})^T$. Note that when B is orthogonal, then $B^* = B$, and when B is diagonal (more generally, symmetric), then $B^* = B^{-1}$. Indeed, the equation satisfied by x in $BH(s, (Q, p, s, X))$,

$$s^T \left(B^{-1}x\right) \leq Q(p, s, X),$$

is equivalent to

$$((B^{-1})^T s)^T x = (B^*s)^T x \leq Q(p, s, X).$$

The norm of s is one, but not necessarily that of B^*s; therefore, we divide both sides by $\|B^*s\|$,

$$(1/\|B^*s\|)(B^*s)^T x \leq (1/\|B^*s\|)Q(p, s, X). \tag{3.5}$$

By the scale equivariance of the quantile operator, and by the relationship $Q(p, s, AX) = Q(p, A^T s, X)$, which follows directly from the definition, we obtain that the right-hand side of (3.5) is equal to

$$Q(p, s, X/\|B^*s\|) = Q(p, s/\|B^*s\|, X) = Q(p, B^*s/\|B^*s\|, BX).$$

Since the transformation $BX + b$ is one-to-one, the transformed intersection of halfspaces is the intersection of transformed halfspaces. Therefore, the transformed directional quantile envelope is by (3.4)

$$\bigcap_{s\in\mathbb{S}^{d-1}} (BH(p, s, X) + b) = \bigcap_{s\in\mathbb{S}^{d-1}} H(p, B^*s/\|B^*s\|, BX + b).$$

The proof is concluded by observing that $s \mapsto B^*s/\|B^*s\|$, where $B^* = (B^{-1})^T$, is a one-to-one transformation of $\mathbb{S}^{d-1}$ onto itself—as can be seen by the direct verification involving its inverse, $t \mapsto B^T t/\|B^T t\|$. □

If we suppress the underlying directional quantile lines (still visible in Figure 3.5), we realize that the plot can accommodate several p simultaneously. Figure 3.6 shows directional quantile envelopes for several p listed there. In the central part, the contours resemble those obtained by fitting normal distribution; in the tail area, they adapt more to the specific shape of the data. Theorem 3.3.4 below shows that the contours of normal distribution are actually directional quantile envelopes. They may be empty; to learn more about this, we can use the close relationship of directional quantile envelopes to a well-known data-analytic concept, halfspace depth, which will be discussed in the next chapter.

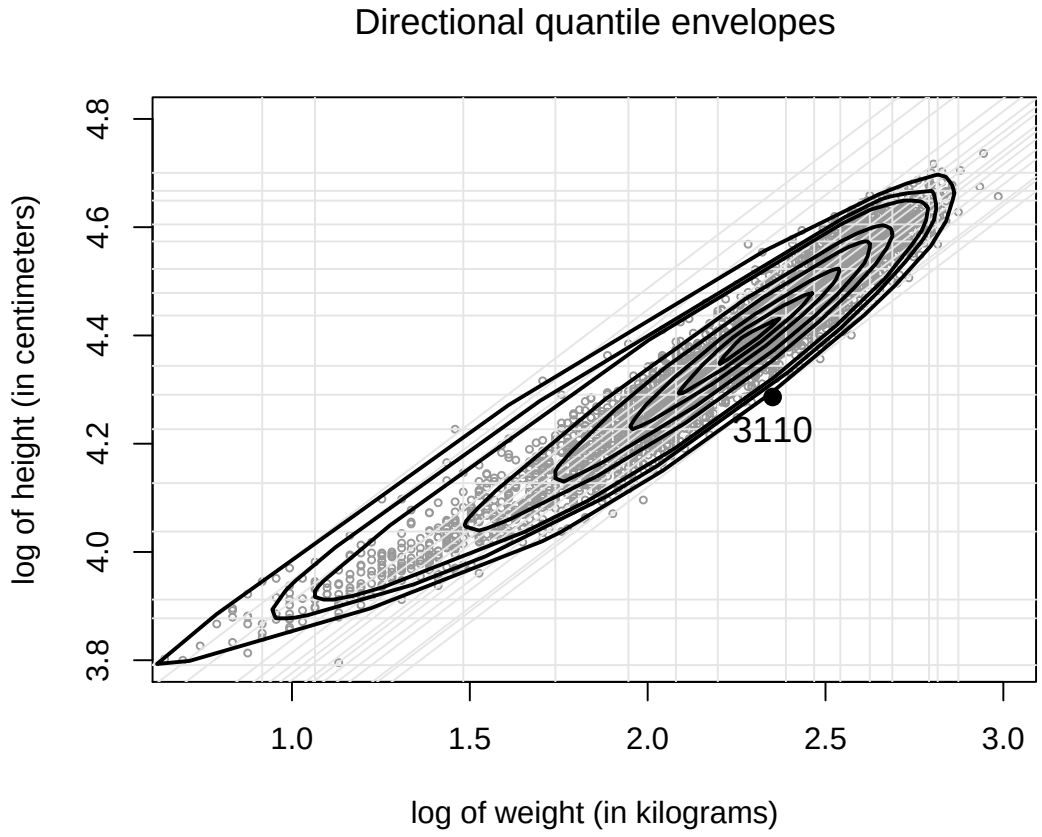

Figure 3.6: *Directional quantile envelopes with different p values, where the index* $p = \{.001, .005, .01, .05, .10, .20, .30, .40\}$. *In the central part, the contours resemble those obtained by fitting normal distribution; in the tail area, they adapt more to the specific shape of the data. The plot can accommodate several* p *simultaneously, and the contours allow for relatively faithful and straightforward retrieval of the directional quantile information.*

Normal distribution plays an important role not only in theoretical study but also in practical applications. However, normal distribution sometimes may be an ideal assumption and may be not be that common in real life. Actually, for many studies, one way to relax the assumption of normality while still retaining some specific structure is via the class of elliptically symmetric distributions. Recall that an *elliptically symmetric*

distribution is the distribution which can be transformed by an affine transformation to a circularly symmetric, rotationally-invariant distribution. Let X be a random vector in $\mathbb{R}^d$ following an elliptically symmetric distribution, if it has density function, it must be in the form

$$|\Sigma|^{-1/2} g\left((X-\mu)^T \Sigma^{-1}(X-\mu)\right), \tag{3.6}$$

where g is a nonnegative scalar function. The following theorem provides the closed form of directional quantile envelopes of a random vector X when it follows an elliptically symmetric distribution with the density function in the form of (3.6).

Theorem 3.3.4 *Suppose that the density function of a random vector X in $\mathbb{R}^d$ is in the form of* (3.6)*. Then it follows an elliptically symmetric distribution, and for any $p \in (0, 1/2)$, the p-th directional quantile envelope is*

$$\{x:\ (x-\mu)^T \Sigma^{-1}(x-\mu) \le c_p^2\}, \tag{3.7}$$

where c_p is a constant such that $\mathbb{P}[s^T \Sigma^{-1/2}(X-\mu) \le c_p] = p$ with $s \in \mathbb{S}^{d-1}$.

Proof. The crucial part of the proof is to realize that $u^T \Sigma^{-1/2}(X-\mu) \sim v^T \Sigma^{-1/2}(X-\mu)$ for any $u, v \in \mathbb{S}^{d-1}$ as $\Sigma^{-1/2}(X-\mu)$ is circularly symmetric. Therefore, for any $s \in \mathbb{S}^{d-1}$, $\mathbb{P}[s^T \Sigma^{-1/2}(X-\mu) \le c_p] = p$. Another fact we need is that $c_p < 0$ as $p \in (0, 1/2)$.

Let $x_0 \in \{x:\ (x-\mu)^T \Sigma^{-1}(x-\mu) \le c_p^2\}$, to show that $x_0 \in D(p)$ we need to prove $\mathbb{P}[s^T X \le s^T x_0] \ge p$ for any $s \in \mathbb{S}^{d-1}$. Notice that $s^T X \le s^T x_0$ is equivalent to

$$(\Sigma^{1/2} s)^T / \|\Sigma^{1/2} s\| \Sigma^{-1/2}(X-\mu) \le (\Sigma^{1/2} s)^T / \|\Sigma^{1/2} s\| \Sigma^{-1/2}(x_0-\mu), \tag{3.8}$$

$(\Sigma^{1/2} s)^T / \|\Sigma^{1/2} s\| \in \mathbb{S}^{d-1}$, and $(\Sigma^{1/2} s)^T / \|\Sigma^{1/2} s\| \Sigma^{-1/2}(x_0-\mu)$ is minimized at $s = -\Sigma^{-1}(x_0-\mu)/\|\Sigma^{-1}(x_0-\mu)\|$ with value $-(x_0-\mu)^T \Sigma^{-1}(x_0-\mu)/\|\Sigma^{-1/2}(x_0-\mu)\| = c_p$. Therefore, we have

$$\mathbb{P}[s^T X \le s^T x_0] \ge \mathbb{P}[(\Sigma^{1/2} s)^T / \|\Sigma^{1/2} s\| \Sigma^{-1/2}(X-\mu) \le c_p] = p.$$

As a result, $x_0 \in H(s, Q(p, s))$, and thus $x_0 \in D(p)$.

Conversely, we need to show that if $x_0 \in D(p)$, then $(x_0-\mu)^T\Sigma^{-1}(x_0-\mu) \leq c_p^2$. Let $x_0 \in D(p)$, then form Definition 3.3.1, $\mathbb{P}[s^T X \leq s^T x_0] \geq p$ for any $s \in \mathbb{S}^{d-1}$. Using (3.8) and with similar deduction, we will have

$$\mathbb{P}[(\Sigma^{1/2}s)^T/\|\Sigma^{1/2}s\|\Sigma^{-1/2}(X-\mu) \leq -(x_0-\mu)^T\Sigma^{-1}(x_0-\mu)/\|\Sigma^{-1/2}(x_0-\mu)\|] \geq p.$$

Therefore $-(x_0-\mu)^T\Sigma^{-1}(x_0-\mu)/\|\Sigma^{-1/2}(x_0-\mu)\| \geq c_p$, which means $(x_0-\mu)^T\Sigma^{-1}(x_0-\mu) \leq c_p^2$. That completes the proof. □

Actually, (3.7) is the contour of elliptically symmetric distribution when indexing by the tangent mass. As a matter of fact, Theorem 3.3.4 indicates that the contours of elliptically symmetric distribution are the directional quantile envelopes. Specially, the contours of normal distribution are actually directional quantile envelopes, as normal distribution is elliptically symmetric. Therefore we indeed extend the approach of normal fitting to multivariate data.

Chapter 4

Probabilistic and statistical interpretation of directional quantile envelopes [1]

4.1 Indexing quantile contours

At the end of the last chapter, we proposed an approach using directional quantile envelopes to plot and visualize directional quantiles, however, there is still a crucial technical question. In this section, we will clarify this technical question, that of indexing: which particular contours or envelopes should correspond to which p? To simplify the question and better elucidate the concept, we will demonstrate it on multivariate quantile contours via normal distribution fitting, as symptomatically done by Evans (1982). As we have already discussed, the fitted contours of normal distribution indexed by p and $(1-p)$ are bound to coincide—normal distribution is continuous and supported by the whole plane. Nonetheless, we still need to assign contours to p in $(0, 1/2]$, and there are essentially two ways of doing that: indexing by the enclosed mass or indexing by the tangent mass.

[1] A condensed version of Chapters 3-5, co-authored with Dr. Ivan Mizera, has been submitted for publication (arXiv:0805.0056v1, 2008).

4.1.1 Indexing by the enclosed mass

The way of indexing by the enclosed mass extrapolates the univariate fact that the p-th and $(1-p)$-th quantiles together leave $2p$ of the distribution mass outside their convex hull and enclose $1-2p$ inside. For example, the contours corresponding to quantiles $p = \{.05, .10, .20, .30, .40, .50\}$ are those enclosing .90, .80, .60, .40 and .20 of the mass of the fitted normal distribution, together with the contour consisting of the single point located at the mode. The actual numerical values can be determined by a transformation to the standard bivariate normal distribution and the subsequent use of the Rayleigh distribution—the distribution of the norm of two uncorrelated normal distributions with equal variances, also a χ distribution with two degrees of freedom.

The results can be seen in Figure 4.1, where marginal and directional quantiles (with respect to BMI) are plotted in grey lines with $p = \{.10, .20, .30, .40, .50, .60, .70, .80, .90\}$. From the interpretational point of view, we are able to observe that the subject represented by the point 3110 lies in the outstanding 10% of the sample; however, this exceptionality is somewhat "generic"—expressed not only through the company of similar subjects with large weight given the height, but also by the company of those with small weight given the height, of those with small height and weight altogether combined, and of those with large height and weight combined.

Note the striking discrepancy between the marginal quantiles and fitted normal contours. Of course, we do not expect the latter to match the former exactly—after all, their constructions follow different principles, and while normal distribution fits the data perhaps not that badly, this fit is not ideal. Nevertheless, there would be no match even if normal distribution would be a perfect fit. The same situations happen to the directional quantiles—the fitted normal distribution do not interact well with directional quantiles, see Figure 4.1 for the BMI direction.

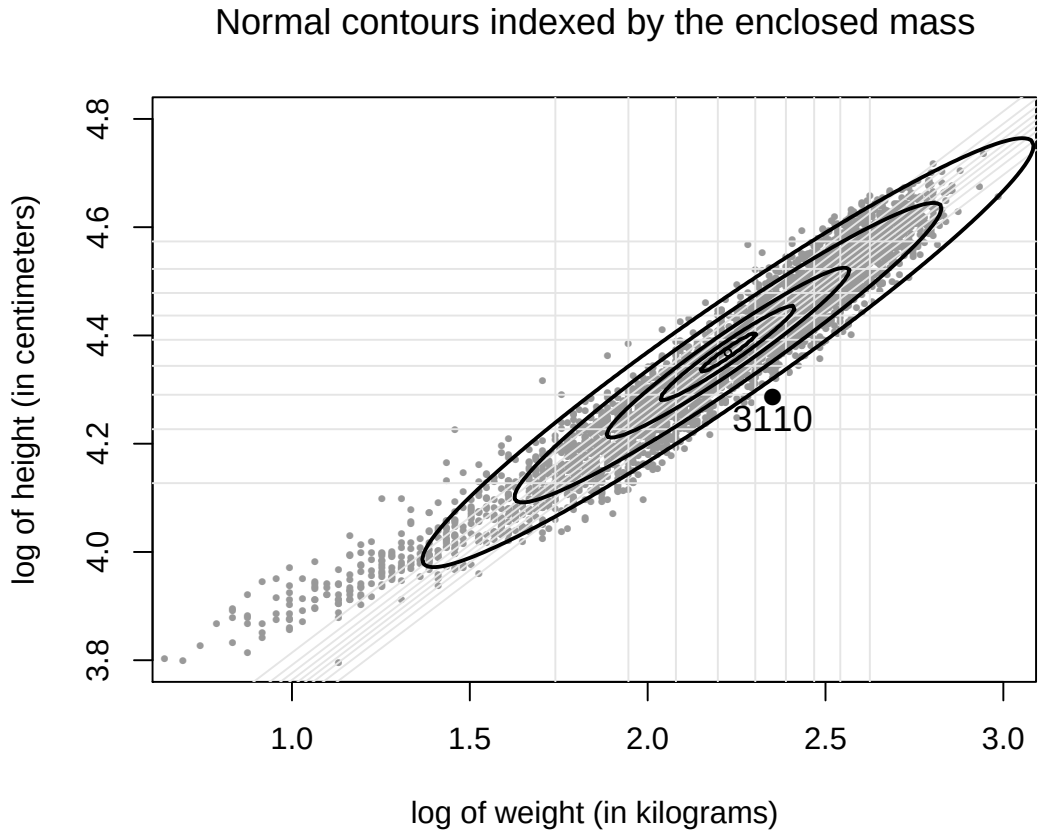

Figure 4.1: *If indexed by the enclosed mass, the contours of the fitted normal distribution do not interact well with directional and marginal quantiles, where in this figure the contour index* $p = \{.20, .40, .60, .80, .90\}$*, respectively.*

4.1.2 Indexing by the tangent mass

An alternative way of indexing is by the tangent mass, extrapolating the univariate fact that both the p-th and $(1-p)$-th quantiles mark the boundaries of the supporting halfspaces containing exactly p of the distribution mass. In the case of multivariate normal distribution, the contour corresponding to p is matching the univariate quantiles indexed by p and $(1-p)$ when projected on any normalized direction, and can be

found by transforming it to the standard normal distribution and the subsequent inverse transformation.

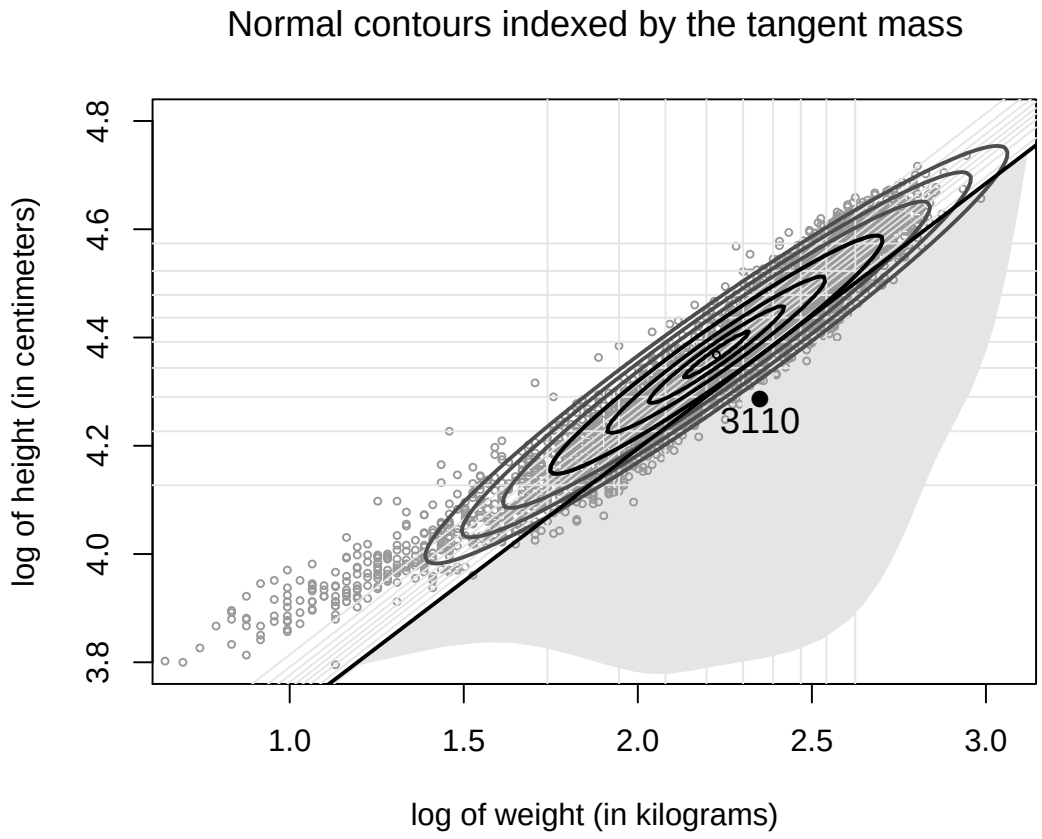

Figure 4.2: *If indexed by the tangent mass, the contours of the fitted normal distribution theoretically match projected quantiles. The halfplane tangent to the contour and passing through the point contains exactly p of the mass of the fitted multivariate normal distribution. Fitting normal contours has all virtues of the ideal, if the data "follow normal distribution"; here such compatibility appears to occur in the central part of the data, but not that much in tail areas, as demonstrated by the additional contours.*

The contours constructed in this way can be seen in Figure 4.2. The grey shaded area illustrates the following property: given any point of the p-th contour, the halfplane passing through the point and tangent to the contour contains exactly p of the mass of

the fitted multivariate normal distribution. The 10% extremality of subject 3110 can be interpreted not as "generic" now, but "substantial": it is given by the company of subjects with similar nature, those with large weight given the height. Note that the boundary of the greyed halfspace is almost identical to the line indicating the (0.9)-th quantile of the *BMI*; hence the picture shows that in this case, the extremality of subject 3110 may be interpreted in terms of *BMI*. Those who are interested in other linear combinations may find it works as well.

Contrary to the indexing by the enclosed mass, indexing by the tangent mass interacts well with marginal and directional quantiles. Even though the match in Figure 4.2 is not perfect—as obviously normal distribution fitting is not the best choice after all—it is much better than that in Figure 4.1. If the data follow normal distribution, then any p-th directional quantile line can be retrieved as a line tangent to the p-th contour in the given direction.

4.1.3 Directional quantile envelopes indexed by tangent mass

Obviously, an approach based on fitting a parametric family of distributions is productive only if the data "follow that distribution"—that is, if the parametric family adapts well to the features present in the data. In an infrequent situation when this is not the case, an often attempted rescue is some ad hoc transformation to the canonical situation, for example, Box-Cox transformation in univariate case (Pere, 2000). Wei, Pere, Koenker and He (2006) demonstrated the pitfalls of such engineering, by showing that the results may be difficult to interpreted in practice; more importantly, it may obscure features that can be unveiled by a more principled nonparametric methodology.

Otherwise, the approach through fitting normal contours could be considered ideal: the contours are a very good summary of the projected quantiles. After all, methods based on normality constitute the core of classical statistics—and despite all dissent, its

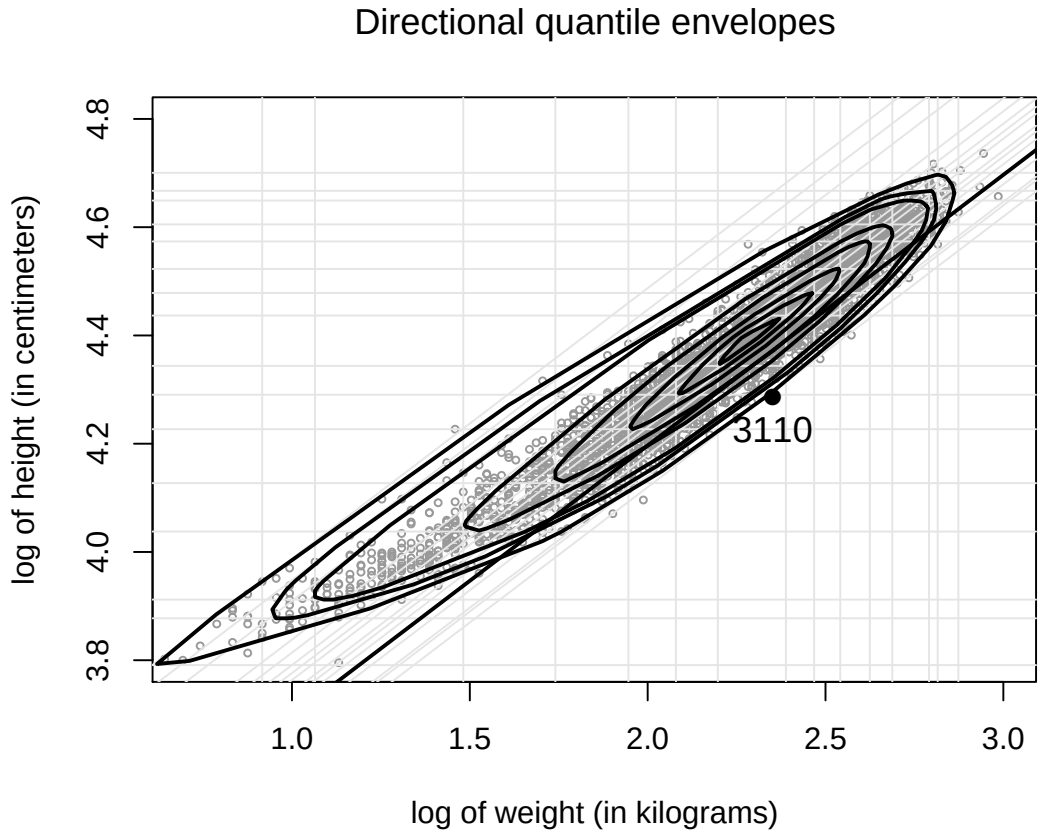

Figure 4.3: *Directional quantile envelopes indexed by the tangent mass interact well with the directional and marginal quantiles. The black line is the directional quantile line in the BMI direction with* $p = .10$*. The envelope indices are* $p = \{.005, .001, .01, .05, .10, .20, .30, .40\}$.

applied core. It is not surprising that this way of thinking is usually the first (if not last) resort of practitioners; in the biometric context, the approaches to "reference contours" related somehow to normal distribution were pursued by Fatti, Senaoana and Thompson (1998) and Pere (2000).

Therefore, our—revised—ideal is to construct an alternative that would be nonparametric, but would extend the normal fitting approach. In other words, the fitting

will coincide with the normal fitting in the case that the data follow normal distribution, as shown in Theorem 3.3.4. We prefer indexing by the tangent mass, because it interacts much better with the directional and marginal quantile information than any other indexing convention, if not only indexing by the enclosed mass. Finally, note that the indexing calculations used transformations to and from the standard normal, for which the multivariate normal family is well suited, due to its *affine equivariance*.

In Figure 4.3, it is shown that the directional quantile envelopes indexed by tangent mass with $p = \{.005, .001, .01, .05, .10, .20, .30, .40\}$. In the central part, the contours resemble those obtained by fitting normal distribution comparing to Figure 4.2; in the tail area, they adapt more to the specific shape of the data. The tangent spaces in the BMI directions match the directional quantile lines well. The plot can accommodate several p simultaneously, and the contours allow for relatively faithful and straightforward retrieval of the directional quantile information—this will be addressed in the following sections. If indexed by tangent mass, Theorem 3.3.4 shows that the contours of elliptically symmetric distribution are the directional envelopes. Specially, the contours of normal distribution are actually directional quantile envelopes, as normal distribution is elliptically symmetric. Therefore we really do extend the approach of normal fitting through a nonparametric methodology.

4.2 Connection to halfspace depth

Except for the fact that the directional quantile envelope approach is an extension of normal distribution fitting, as verified in Theorem 3.3.4, there is a close relationship between directional quantile envelopes and the well-known data-analytic concept, depth. This makes our directional approach more appealing. The depth is halfspace depth, also known as Tukey depth reflecting the fact that Tukey (1975) proposed halfspace

depth contours for plotting bivariate data even though it was Hodges (1955) who first introduced it.

Let P be the distribution of a random variable X in $\mathbb{R}^d$. The *halfspace depth*, $d(x)$, of a point $x \in \mathbb{R}^d$, is defined as $\inf P(H)$, where H runs over all closed halfspaces containing x (or, equivalently, over all closed halfspaces with x lying on their boundary). The p-th upper level set of depth is $\{x : d(x) \geq p\}$ and the p-th lower level set of depth is $\{x : d(x) > p\}$. The upper level sets of depth is also known as halfspace depth contours, denoted by D_p for $0 < p < 1$.

Theorem 4.2.1 *For every $p \in (0, 1/2]$, the p-th directional quantile envelope is equal to the p-th upper level set of depth: $D(p) = D_p$.*

Proof. If $x \in D(p)$, recalling Definition 3.3.1 of $D(p)$, $D(p) = \bigcap_{s \in \mathbb{S}^{d-1}} H(s, Q(p, s))$, then we have $x \in H(s, Q(p, s))$ for every $s \in \mathbb{S}^{d-1}$. Thus $P(\{y : s^T y \geq s^T x\}) \geq p$ for all $s \in \mathbb{S}^{d-1}$ as a direct consequence of Definition 1.1.1 of $Q(p, s)$, which means $d(x) \geq p$. That is, $x \in D_p$.

Conversely, if $x \in D_p$, that is, x satisfies that $d(x) \geq p$, then for every $s \in \mathbb{S}^{d-1}$ we have $P(\{y : s^T y \geq s^T x\}) \geq d(x) \geq p$. It follows, in view of $\inf$ in Definition 1.1.1, that $s^T x \geq Q(p, s)$ and thus $x \in H(s, Q(p, s))$. Hence $x \in D(p)$ from the definition of $D(p)$, which completes the proof. □

Theorem 4.2.1 is true only for the "inf" quantile version defined by Definition 1.1.1. Any other quantile version gives smaller envelopes and hence are included by the corresponding halfspace depth contours. More precisely, for every suitable p (less than the depth of Tukey's median), the p-th directional quantile envelope given by any quantile version includes $\{x : d(x) > p\}$, the lower level set of depth, and is included by $\{x : d(x) \geq p\}$, the upper level set of depth. Obviously, in some situations,

for instance, the distribution P of a random variable X is absolutely continuous with positive densities, the directional quantile envelopes given by any version of quantiles are the same as the corresponding halfspace depth contour, when the lower level set is equal to the upper level set.

All interpolated versions of quantiles yield somewhat smaller envelopes; Rousseeuw and Ruts (1999) point out that this is also the case for the related notion of halfspace trimmed contours of Massé and Theodorescu (1994). These subtle differences vanish in regular situations as well—for instance, for absolutely continuous distributions with positive densities, the same case as in directional quantile envelopes.

Theorem 4.2.1 implies that directional quantile envelopes are nonempty for $p \leq 1/(d+1)$, in the two-dimensional case for $p \leq 1/3$, due to a basic result from depth theory known as a centerpoint theorem, the existence of centerpoint (the parameter with at least $1/(d+1)$ depth), see Donoho and Gasko (1992) and Mizera (2002). More generally, directional quantile envelopes are nonempty for any $p \leq \max\{d(x) : x \in \mathbb{R}^d\}$. In the case of continuous distributions, the greatest possible value is $.5$, for instance if P is any elliptically symmetric distribution. In the case of discrete data points, the depth can be as high as 1—when all the mass is shrinking at a point, less interested though in practice. However, it is worth noting that technically the value of halfspace depth can be any value between 0 and 1.

One reason we refer to what are essentially depth contours as "directional quantile envelopes", is the existence of various interpolated quantile versions—which we would prefer in practical use, in particular because they allow for construction of contours interpolating between various depth level sets. Another reason, and maybe most important, is the possibility of adding covariates to the construction of directional quantile envelopes, for instance, bivariate growth charts, which will be discussed in the

next chapter. Also, while Theorem 4.2.1 is rigorously true only for the "inf" quantile version following Definition 1.1.1, most of the other theorems presented in this thesis are true also for other versions—the detailed discussions are listed at the end of every proof of the theorems.

Except for the consequence of the centerpoint theorem, it turns out that there is not much existing depth theory contributing to quantile tomography as directional quantile philosophy may shed light on depth. Mathematically, we work with objects dual to convex sets, the support functions, as defined in convex analysis context. In this way, we may continue the line of thought brought into statistics by Walther (1997a,b).

Recall that the support function of a convex set K in $\mathbb{R}^d$ is defined as $\sigma_K(u) = \sup_{s\in\mathbb{S}^{d-1}} s^T u$, where $\mathbb{S}^{d-1} = \{s : \|s\| = 1, s \in \mathbb{R}^d\}$ as before. Every support function is *positively homogeneous*, $\sigma_K(cu) = c\sigma_K(u)$ for every u and every $c > 0$, and *sublinear*, $\sigma_K(u+v) \leq \sigma_K(u)+\sigma_K(v)$. Conversely, every positively homogeneous and sublinear function is the support function of some convex set. See Rockafellar (1996) for more details.

The equivariance of directional quantiles means that, for fixed p, the directional quantile function $q(u) = Q(p, u^T X)$ is positively homogeneous, easy to be verified; nevertheless, it is not hard to find an example when it is not sublinear. We can consider the maximal support function, that is, the maximal positively homogeneous and sublinear function among those dominated by the directional quantile function. It turns out that the result is the support function of the p-th directional quantile envelope. While this observation is not directly helpful in its technical aspect, it can be effectively used in the algorithm constructing bivariate directional quantile envelopes, which will be discussed in a following chapter.

4.3 Probabilistic information and retrieval theorem

Directional quantile lines are very important, from which we can extract the directional quantile information: how quantile lines divide the data. A crucial question is how far is it possible to recover the directional quantile lines from the directional quantile envelopes—what is the price for the compression of the directional quantile information? This question will not be observed from the aspect of depth contours but rather from the view of directional quantile envelopes. Although the general theory is possible in arbitrary dimension, in what follows we will limit ourselves to a two dimensional case, to avoid unnecessary complex notation.

Let e be a point lying on the boundary, ∂E, of a bounded convex set $E \subset \mathbb{R}^d$. A *tangent* of E at e is any hyperplane (line) containing e that has empty intersection with the interior of E. Such a line determines the corresponding *tangent halfspace*, the halfspace that has the tangent as its boundary and its interior does not contain any point of E. Mathematically, if $\{x : s^T x = s^T e\}$ is the tangent of E passing through the point e and $s^T x < s^T e$ holds for any x belongs to the interior of E where $s \in \mathbb{S}^{d-1}$, then $\{x : s^T x \geq s^T e\}$ is the tangent halfspace passing through the point e; obviously $\{x : s^T x \geq s^T e\}$ has no intersection with the interior of E.

Let X be a random vector with distribution P in $\mathbb{R}^d$. The *maximal mass at a hyperplane* is defined as

$$\Delta(P) = \sup \left\{ \mathbb{P}[s^T X = c] : s \in \mathbb{S}^{d-1}, c \in \mathbb{R} \right\}.$$

In a one dimensional case, P is continuous is equivalent to $\Delta(P) = 0$, no point has positive probability mass. In the situation that P is an empirical distribution with finite sample size n, $\Delta(P)$ is at least $1/n$. The following theorem is essential in interpreting directional quantile envelopes when retrieving probabilistic information from them.

Theorem 4.3.1 *Let P be the distribution of a random variable X in $\mathbb{R}^d$, and let $p \in (0, 1/2]$. If H is a tangent halfspace of $D(p)$, then $p \leq P(H) \leq 2p + \Delta(P)$. Moreover, $p \leq P(H) \leq p + \Delta(P)$, if ∂H is the unique tangent of $D(p)$ at some point from $H \cap \partial D(p)$; in particular, $P(H) = p$ if $\Delta(P) = 0$.*

If $A \in \mathbb{S}^{d-1}$ is a finite set of directions and H is a tangent halfspace of $D_A(p)$, then still $P(H) \leq 2p + \Delta(P)$, and $p \leq P(H) \leq p + \Delta(P)$, if ∂H is the unique tangent of $D_A(p)$ at some *point from $H \cap \partial D(p)$. In particular, $P(H) = p$ if $\Delta(P) = 0$.*

Proof. Suppose that $y \in \partial D_A(p)$; we claim that there is a direction $s \in A$ such that $y \in \partial H(p, s)$. Otherwise, there exists an $\epsilon > 0$ such that $s^T y - Q(p, s) \geq \epsilon$ for any $s \in A$. The ball centered at y with radius $\epsilon/2$ would then belong to $H(p, s)$ for any $s \in A$ and thus belong to $D_A(p)$ as well from Definition 3.3.1, that means, y is an interior point of $D_A(p)$, a contradiction.

For $A = \mathbb{S}^{d-1}$, the inequality $p \leq P(H)$ follows from the fact, established by Theorem 4.2.1, that the depth of every $y \in D(p)$ is at least p. From Definition 2.5.1 on halfspace depth, if $d(y) \geq Q(p, s)$, then any halfspace containing y has at least probability mass p. That is, $p \leq P(H)$.

Suppose now that ∂H is a unique tangent of $D_A(p)$ at y, then its direction is $s \in A$, and $H = H(p, s)$. Otherwise, let us denote the direction of H by t. The uniqueness of the tangent means that any other directional quantile hyperplane passing through y in any direction $u \in A$, $u \neq t$, has a common point with the interior of $D_A(p)$. Particularly, we pick $u = s$ and denote the common point by x. An ϵ-ball centered at such a point x is contained in $H(p, s)$ as the ϵ-ball is contained by $D_A(p)$, hence $\partial H(p, s)$ is parallel to the directional quantile hyperplane $\{x : s^T x = s^T y\}$ in distance more than $\epsilon > 0$, and y lies in the interior of $H(p, s)$, which is contradictory to the fact that $y \in \partial H(p, s)$. Therefore, $H = H(p, s) = \{x : s^T x \leq Q(p, s)\}$, and we have $P(H) \geq p$ recalling the

definition of $Q(p, s)$.

On the other hand, if X is a random vector with distribution P, we have

$$P(H) = \mathbb{P}[s^T X \leq Q(p, s) - \epsilon_n] + \mathbb{P}[Q(p, s) - \epsilon_n \leq s^T X \leq Q(p, s)],$$

where $\mathbb{P}[s^T X \leq Q(p, s) - \epsilon_n] < p$ and $\mathbb{P}[Q(p, s) - \epsilon_n \leq s^T X \leq Q(p, s)] \leq \Delta(P)$ as $\epsilon_n \to 0$. That means that $P(H) \leq p + \Delta(P)$.

It remains to prove the inequality $P(H) \leq 2p + \Delta(P)$ when the tangent H is not unique. For simplicity, we assume that $d = 2$, the argument for $d > 2$ being similar. Let $T_{D_A(p)}(y)$ be the directions of all tangents of the convex set $D_A(p)$ at the point y, then it is a convex cone, a pie slice in two dimensional case. Suppose u is one endpoint of the closed segment $T_{D_A(p)}(y) \cap \mathbb{S}^1$. Theorem 24.1 of Rockafellar (1996) implies that there is a sequence $y_n \in \partial D_A(p)$ such that $y_n \to y$ and y_n has a unique tangent with direction s_n such that $s_n \to u$. The convergence $s_n^T X \to u^T X$ and $s_n^T y_n \to u^T y$ imply that $\liminf_{n\to\infty} \mathbb{P}[s_n^T X < s_n^T y_n] \geq \mathbb{P}[u^T X < u^T y]$. Using the fact proved in the first part of the proof, we obtain $s_n^T y_n = Q(p, s_n)$, because y_n has unique tangent; this implies $\mathbb{P}[s_n^T X < s_n^T y_n] < p$ and thus $\mathbb{P}[u^T X < u^T y] \leq p$. If v is the other endpoint of $T_{D_A(p)}(y) \cap \mathbb{S}^1$, then we have $\mathbb{P}[v^T X < v^T y] \leq p$. Finally, $t \in T_{D_A(p)}(y)$ implies that $H \subseteq \{x : u^T x < u^T y\} \cap \{x : v^T x < v^T y\} \cap \{y\}$ and thus

$$P(H) \leq P(\{x : u^T x < u^T y\}) + P(\{x : v^T x < v^T y\}) + P(\{y\}) \leq 2p + \Delta(P). \square$$

Theorem 4.3.1 provides a practical guideline for the recovery of the directional quantile information. The left panel of Figure 4.4 shows the situation when the tangent to the directional quantile envelope is unique. For a population distribution with $\Delta(P) = 0$, we can uniquely identify the directional quantile in the direction perpendicular to the tangent pointing outside. The boundary of the tangential halfspace is the p-th directional quantile line, in the given direction; and hence has probability mass p. The visual

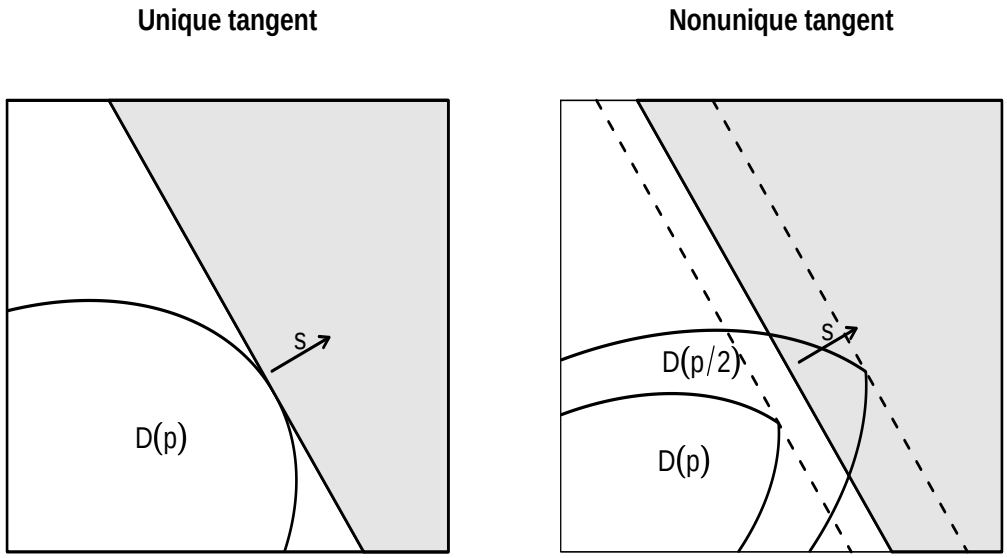

Figure 4.4: *Left panel: if the tangent line to the p-th directional quantile envelope is unique, then the tangential halfspace is the p-th directional quantile halfspace, in the given direction. Right panel: if the tangent line is nonunique, then this directional quantile halfspace lies between p-th and $(p/2)$-th directional quantile envelope.*

determination of such a uniqueness may be slightly in the eye of beholder; if this is not desirable, we may switch to a strictly finite-sample viewpoint, in which the directional quantile envelopes of empirical probability distributions are polygons and the uniquely identifiable directional quantile lines are those that contain a boundary segment of the polygon. Such a structure can always be discovered under appropriate magnification of the polygon. If the tangent is not unique, the situation shown in the right panel of Figure 4.4, then the exact identification of the directional quantile line is not possible; nevertheless, the inequality $P(H) \leq 2p$ given by Theorem 4.3.1 when $\Delta(P) = 0$ allows at least for its approximate localization, especially when the plotted quantile envelopes are so chosen that p follows a geometric progression with multiplier $1/2$ (as in Figure 4.4; note that such choice gives approximately equispaced contours for normal

distribution in the tail area).

A boundary point of a convex set that admits more than one tangent is called a *rough* (singular) point, also known as corner point. It is known—see Theorem 2.2.4 of Schneider (1993)—that such points are quite exceptional; in particular, for any closed convex set in $\mathbb{R}^d$, rough points is at most countable. Convex, closed subsets of $\mathbb{R}^d$ having no rough points are called *smooth*, consistently with the natural geometric perception of the boundary in this case. If $D(p)$ is smooth, then the collection of its tangent halfspaces is in one-one correspondence with the collection of p-th directional quantile halfspaces, sharing the same boundaries, but in opposite directions.

Corollary 4.3.1 *If $D(p)$ is smooth, then the boundary hyperplane of its tangent halfspaces is a p-th directional quantile hyperplane on some direction $s \in \mathbb{S}^{d-1}$.*

Proof. If $D(p)$ is smooth, we know that $\Delta(P) = 0$. Given a direction $s \in \mathbb{S}^{d-1}$, from Theorem 4.3.1, there exists a unique tangent halfspace, denoted by H, on the direction s such that $P(H) = p$. Suppose that the boundary hyperplane is $\{x : s^T x = q\}$, then $H = \{x : s^T x \geq q\}$ and thus $P(\{x : s^T x \geq q\}) = p$. We claim that $Q(p, -s) = -q$. To verify our claim, we need to show that $-q = \inf\{u : \mathbb{P}[-s^T X \leq u] \geq p\}$. Otherwise, there exists $-c < -q$, such that $P\{x : -s^T x \leq -c\} \geq p$, which implies that $P\{x : s^T x \geq c\} \geq p$. Therefore, we have $P\{x : q < s^T x < c\}$=0, which is contradictory to the smoothness of $D(p)$. Our claim and thus the corollary hold true. □

Although the assumption of smoothness may sound optimistically mild, the examples in Rousseeuw and Ruts (1999) show that distributions with depth contours having a few rough points are not that uncommon. It may be argued that all these examples have somewhat contrived flavor, especially when the support of the distribution is some regular geometric figure. It is not impossible that typical population distributions have

smooth depth contours—however, we were not able to find a suitable formal condition reinforcing this belief, beyond the somewhat restricted realm of elliptically-contoured distributions. Recall that the distribution P of a random variable is called *elliptic* if it can be transformed by an affine transformation to a circularly symmetric, rotationally-invariant distribution. If the elliptically symmetric distribution P has density, then it must be in the form (3.6). That is,

$$f(x) = |\Sigma|^{-1/2} g\left((x-\mu)^T \Sigma^{-1} (x-\mu)\right).$$

Theorem 4.3.2 *Every elliptically symmetric distribution has smooth directional quantile envelopes $D(p)$, for every $p \in (0, 1/2)$.*

Proof. By rotational invariance, the directional quantile envelopes of any circularly symmetric distribution are circles; since elliptically symmetric distributions are those that can be transformed to the circular symmetric ones by an affine transformation, the theorem follows from their affine equivariance (and holds true for any quantile version). □

As a consequence, we are able to recover all the directional quantile lines for elliptically symmetric distributions. In particular, this confirms the fact mentioned earlier: normal contours allow for the retrieval of *all* directional quantile lines, as normal distribution is elliptically symmetric. For other distributions, this may or may not be true. For instance, we can retrieve all directional quantile lines from the product of a t distribution with degrees of freedom 2 or more with itself; but only marginal quantiles from the product of the Cauchy distribution with itself. For various uniform distributions, we can retrieve many directional quantile lines, but not all.

Of course, the quantile envelope always gives a conservative estimate of the directional quantile; we know that the tangent halfspace, if not unique, may contain more, but in no case less than p of the distribution mass. However, this deduction can

only be made from the nonempty envelope—and it is known that the upper level set of depth is guaranteed nonempty only for $p \leq 1/3$; in the general d-dimensional case, this number is equal to $1/(d+1)$. Apparently, condensing directional quantile information comes for a price; the question how high this price is may need some further insight.

4.4 Consequence for depth characterization problem

Even if the tangent line at a boundary point of a directional quantile envelope is non-unique, it does not necessarily mean that the information about certain directional quantiles is lost. Even if the directional quantile is not retrievable from the envelope *directly*, in a straightforward manner, it may be possible to reconstruct it from the totality of these envelopes. From the formal point of view, it would mean that the collection of directional quantile envelopes, for all $p \in (0, 1/2]$, determines the distribution uniquely (and then, of course, all directional quantiles).

Surprisingly, this plausible property has not yet been rigorously proved in full generality. In the depth context, this is a long standing open problem: Does the depth function uniquely characterize the probability distribution? Positive answers have been established for partial cases: depth functions uniquely characterize empirical distributions (Struyf and Rousseeuw, 1999), and more generally atomic distributions (Koshevoy, 2002; Hassairi and Regaieh, 2007), and also absolutely continuous distributions with compact support (Koshevoy, 2001). Certain progress in this line is also our following result regarding distributions with smooth depth contours, directional quantile envelopes. Note that these include, via Theorem 4.3.2, elliptically symmetric distributions, which may have unbounded support— hence the following theorem is not covered by that of Koshevoy (2001).

Theorem 4.4.1 *If the directional quantile envelopes $D(p)$, of a probability distribution*

P in $\mathbb{R}^d$ with contiguous support, have smooth boundaries for every $p \in (0, 1/2)$, then there is no other probability distribution with the same directional quantile envelopes.

Proof. Without loss of generality, we may assume that the origin is the point with maximal depth, that is the Tukey median. We claim that for any $s \in \mathbb{S}^{d-1}$ and $c \leq 0$,

$$P(\{x : s^T x \leq c\}) = p^*,$$

where $p^* = \sup\{d(x) : s^T x = c\}$, the maximal depth on the hyperplane $\{x : s^T x = c\}$.

If $p^* = 0$, then $d(x) = 0$ for any $x \in \{x : s^T x = c\}$, which means that $D(p) \cap \{x : s^T x = c\} = \emptyset$ for any $p \in (0, 1/2)$. Considering the support of P, *supp(P)* belongs to $\cup_{p\in(0,1/2)} D(p)$, we have $P(\{x : s^T x \leq c\}) = 0$.

If $p^* > 0$, to verify the claim it is sufficient to show that $\{x : s^T x = c\}$ is tangent to $D(p^*)$, which is equivalent to $\{x : s^T x = c\} \cap \partial D(p^*)$ being a singleton as $\partial D(p^*)$ is smooth. Otherwise, $\{x : s^T x = c\} \cap D(p^*)$ would be a compact convex set with nonempty interior; let y be one of the interior points, then $d(y) = p^*$, that is, $P(\{x : t^T x \leq t^T y\}) = p^*$ for some $t \in \mathbb{S}^{d-1}$. As y is also an interior point of $D(p^*)$, there exists a point $z \in D(P^*) \cap \{x : t^T x \leq t^T y\}$ with $t^T(z - y) < 0$. Thus $P(\{x : t^T x \leq t^T z\}) \geq p^*$, which implies $P(\{x : s^T z < t^T x < t^T y\}) = 0$, a contradiction. Therefore, $\{x : s^T x = c\}$ is tangent to $D(p^*)$; and equivalently $\{x : s^T x = c\} \cap D(p^*)$ is a singleton, denoted by y. Obviously $d(y) = p^*$. The fact that any non-tangent halfspace passing through y will contain interior points of $D(p^*)$ implies that $P(\{x : s^T x \leq c\}) = p^*$.

We have proved that for any $s \in \mathbb{S}^{d-1}$, the distribution of $s^T X$ is uniquely determined by $D(p)$; the theorem follows. □

As an example, elliptically symmetric distributions with densities have smooth directional quantile envelopes for every $p \in (0, 1/2)$; and hence they are uniquely

characterized by the directional quantile envelopes. Specially, normal distributions are characterized by the envelopes.

Chapter 5

Estimation and Approximation [1]

5.1 General setting of estimation and equivariance

So far, our quantile deliberations have been rather unsophisticated, from statistical perspective. Since our discussion involved also some data examples, people might get an impression that what we have already proposed is a statistical methodology, that is, some algorithm(s) which can be used for processing the data. In fact, this has yet to be done; it is important to realize that our considerations up to now were more probabilistic than statistical in spirit. That is, there is still a gap in the application of this method in practice. In this section, we will try to fill the gap and make it possible to address the real data using the method.

Applying what was defined for a general distribution to an empirical distribution indeed means some statistical advancement, in the spirit of the principle of the approach called "naïve statistics" (Hajek and Vorlickova, 1977) or "analogy" (Goldberger, 1968; Manski, 1988) or "plug-in" (Efron and Tibshirani, 1993): the result of the evaluation of a functional on the population distribution is estimated via the application of the same functional to the empirical distribution supported by the data. While this may be a way of obtaining satisfactory estimates (in fact, this is the exclusive approach considered

[1] A condensed version of Chapters 3-5, co-authored with Dr. Ivan Mizera, has been submitted for publication (arXiv:0805.0056v1, 2008).

in the literature concerning depth so far) there are situations calling for more refined approaches.

Let us outline some general principles. We are interested in the *population quantile information*, that is, directional quantiles of some *population distribution*; we believe that our data comes, in some sampling manner, from this distribution. To facilitate theoretical analysis of typical cases, it is often reasonable to posit some assumptions on the population distribution; while a membership in a parametric distribution family, or elliptically symmetric distribution family may be considered too stringent, continuity assumptions are often acceptable. We may also want to adopt these, but certainly not too stringent ones, like the last two.

The source from which the information is *estimated* is data. Our general strategy is, for fixed p, to estimate the directional quantiles $Q(p, s)$ by $\hat{Q}(p, s)$, and then use these estimates to generate the estimated directional quantile envelope. From the point of view of quantiles, we can obtain more useful and interpretable information to explore the population.

By Theorem 4.2.1, the estimated and population directional quantile envelopes are the upper level sets of depth applied to the empirical and population distributions, respectively. It is known that depth is affine invariant, and thus its upper level sets are affine equivariant as well. Thus we estimate the population depth function by the empirical depth function—the exclusive approach discussed and studies in the depth literature, from which is known that both estimated and population level sets are affine equivariant. In Theorem 3.3.3, we have proved the affine equivariance of population quantile envelopes. The next step is to explore that of estimated quantile envelopes.

In some other situations directional quantiles by some other means, for instance as a response of a quantile regression, the affine equivariance of the resulting envelopes

is not that clear. If the estimates exhibit some form of convergence to population depth contours, then one would have such an equivariance at least approximately due to Theorem 5.4.2; nevertheless, *exact* equivariance holds under mild assumptions on the directional quantile estimators. The following Theorem 5.1.1 is about estimated quantile envelopes, which can be viewed as a generalization of Theorem 3.3.3, the affine equivariance of population quantile envelopes.

Theorem 5.1.1 *Suppose that directional quantile estimators $\hat{Q}(p, s)$ are translation and scale equivariant, for all $s \in \mathbb{S}^{d-1}$ and fixed p. Then the directional quantile envelope generated by these estimators is affine equivariant.*

The proof is the same as that in Theorem 3.3.3 and will not be provided here. For any other quantile version, we need to go through a similar proof procedure to verify the affine equivariance. That is, Theorem 5.1.1 holds true for all quantile versions. In summary, the estimated and population directional quantile envelopes are affine equivariant for all quantile versions.

5.2 An application to extreme quantiles

The case of extreme quantiles is the one where the need for other than empirical estimators of population quantiles is noticeably demonstrated. Assuming 100 observations are available, then their maximum, the p-th empirical quantile for any $p > 0.99$, may not be found satisfactory for estimating a threshold with exceedance probability less than, say, 0.001. In many diverse applications the extreme quantiles of a process are of primary importance. Internet traffic, rainfall, material strength, pollutant concentrations and insurance claim sizes are all examples of process of this type. In all these cases, an estimate, rather than the maxima of the data and usually greater than the quantile index of the maxima, is expected. We will review the two basic approaches for

estimating extreme quantiles in univariate case and extend them to multivariate through our directional quantile approach.

It is worth briefly describing the two basic approaches for estimating extreme quantiles before discussing this topic more thoroughly. The first one is the block maxima approach, which is based on the class of non-degenerate limiting distributions for the maximum under linear normalization. Suppose that $\{x_1, x_2, \cdots, x_n\}$ is a sample from the distribution P of a random variable X; denote its cumulative function by F and the maximum of $\{x_1, x_2, \cdots, x_n\}$ by M_n. The block maxima approach fits the distribution of M_n by the distribution class in the parametric form of

$$G(x) = \exp\left(-\left\{1 + \frac{\gamma(x-\mu)}{\sigma}\right\}_+^{-1/\gamma}\right), \tag{5.1}$$

where $\mu, \sigma > 0$ and γ are the location, scale and shape parameters, respectively, and $z_+ = \max(z, 0)$. The distribution class of G is known as the generalized extreme value distribution: the Gumbel distributions are obtained in the limit as $\gamma \to 0$; the Frechet distributions are obtained for $\gamma > 0$ and the Weibull distributions are obtained for $\gamma < 0$. For large N, the generalized extreme value distribution approximates the distribution of M_N. Given a dataset of maxima, this approximation can be treated as exact, and hence the distribution can be fitted to the data by estimating the parameters (μ, σ, γ). The commonly known block maxima approach can be employed after dividing the sample $\{x_1, x_2, \cdots, x_n\}$ into blocks and considering the maxima derived from each block. The estimation can be performed by maximum likelihood, and likelihood inferential methods can subsequently be used. Among all the alternative estimation methods, the probability weighted moments method of Hosking, Wallis and Wood (1985) is perhaps the most popular one. Extreme quantiles of the distribution F can be estimated similarly on the basis that F_n approximates F.

The second approach is the peaks over threshold (POT) method, which is based on an

approximation to the conditional exceedance probability $\mathbb{P}[X > x|X > u]$ of a random variable X for all x greater than some predetermined large threshold u. Specially, we use the generalized Pareto distribution (Pickands, 1975), conventionally shortened as GPD, to approximate the conditional distribution of exceedances. The 3-parameter generalized Pareto distribution has a probability density function in the form of

$$G(x) = \frac{1}{\sigma}\left(1 - \frac{\gamma(x-\mu)}{\sigma}\right)^{1/\gamma - 1}, \tag{5.2}$$

where μ, σ, and γ are location, scale and shape parameters, respectively. The 3-parameter generalized Pareto distribution represents a large group of distributions. For instance, the Pareto distribution is obtained for $\gamma < 0$; the exponential distribution is obtained for $\gamma \to 0$ and $\mu = 0$. The 3-parameter GPD reduces to a 2-parameter GPD for $\mu = 0$. For statistical inference the 3-parameter GPD is often treated as the exact distribution of the conditional exceedance, and hence the data that exceed u can be used to estimate the generalized Pareto parameters (μ, σ_u, γ). Maximum likelihood or probability weighted moments method can again be applied in estimating. The probability $\mathbb{P}[X > u]$ can be estimated using the empirical proportion of exceedances of u, and the extreme quantiles can be estimated by simply inverting the corresponding model for the unconditional tail probability $\mathbb{P}[X > x]$. It should be pointed out that the generalized Pareto scale parameter σ_u depends on the threshold u. The choice of the threshold is crucial and one common method is to choose the threshold as the lowest value so that the estimates of the 3 parameters are roughly constant.

The above review of the two basic approaches follows Stephenson and Gilleland (2006). Various other approaches to dealing with extreme quantiles in univariate settings can be found in the books of Beirlant, Goegebeur, Teugels and Segers (2004), Reiss and Thomas (2007), Resnick (2007), and the references given there. In multivariate data analysis, the extended block maxima approach is based on the multivariate extreme

value distribution, a generalization of the limiting distribution (5.1), which focuses on componentwise maxima, see the book of Beirlant, Goegebeur, Teugels and Segers (2004). The multivariate extension of the peaks over threshold (POT) approach is not so obvious, although some approaches have been proposed (Coles and Tawn, 1991; and Ledford and Tawn, 1996). However, in our estimating proposal, which is based on the concept of directional quantiles, any one dimensional approach can be extended to multidimension besides the block maxima and the peaks over threshold approaches.

Suppose $\{x_1, x_2, \cdots, x_n\}$ is a sample of the distribution P of a random X in $\mathbb{R}^d$, given a small $p \in (0, 1/2)$, the estimation of the p-th extreme quantile in any direction $s \in \mathbb{S}^{d-1}$ can be obtained by any previously proposed approach, the block maxima approach, the peaks over threshold approach and so on, in terms of the one dimensional data $\{s^T x_1, s^T x_2, \cdots, s^T x_n\}$. The p-th extreme quantile contour of the data set $\{x_1, x_2, \cdots, x_n\}$ then would be obtained by the extreme quantile estimation in all directions. Any univariate extreme quantile approach is eligible for our estimating scheme; the directional extreme quantile envelopes inherit many good properties from their one dimensional counterpart. For example, the estimated directional extreme quantile envelopes are affine equivariant, a desirable property, if the one dimensional estimators are translation and scale equivariant due to Theorem 5.1.1. Both of the two basic approaches are translation and scale equivariant and hence the estimated extreme quantile envelopes based on the two approaches are affine equivariant. The directional extreme quantile envelopes are convex and nested if the corresponding univariate analogs are monotone.

The computation of multidimensional extreme quantile envelopes can directly adopt the algorithm of the corresponding univariate extreme quantiles. Nowadays many one dimensional extreme quantile methods have been implemented in various software,

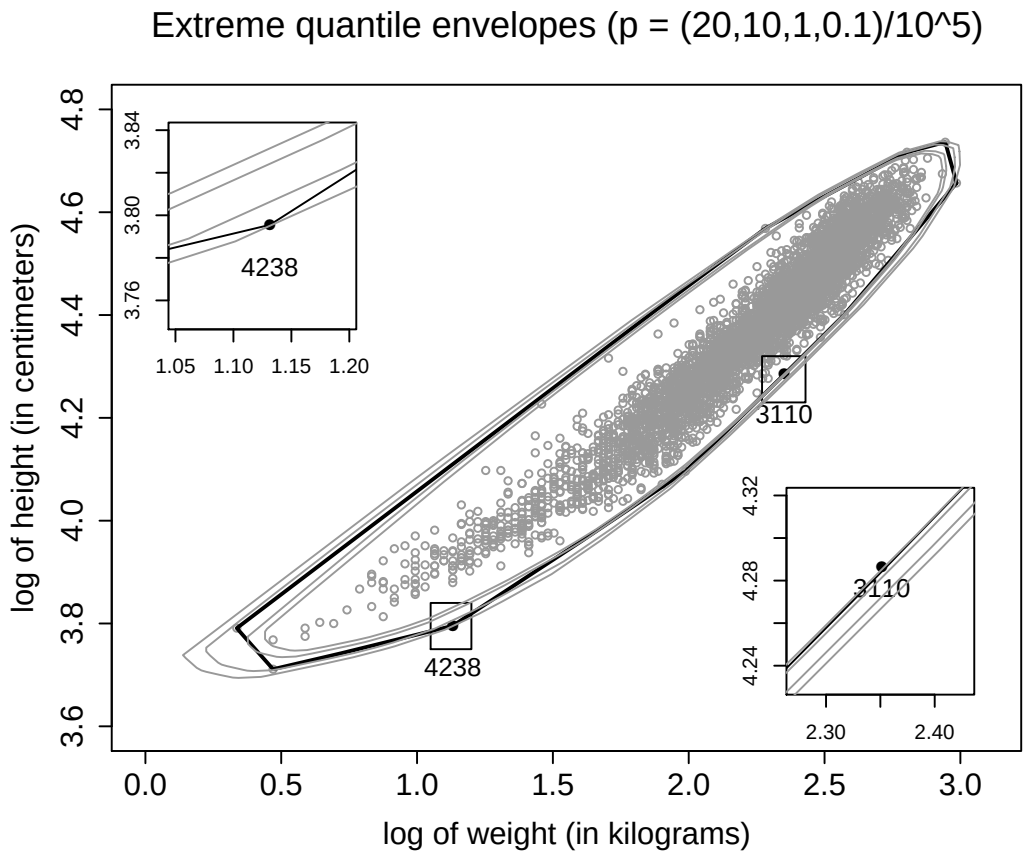

Figure 5.1: *Directional extremal quantiles, derived from the corresponding univariate analogs, and the convex hull, the empirical extremal quantile.*

while only a few provide limited approaches to multivariate extreme quantiles. We do not have a particular preference for any of the methods proposed in the literature on dimension one, and, being focused on the illustration of quantiles for multivariate data. We rather opportunistically chose from those with a more nonparametric flavor, the R package `evir` (McNeil and Stephens, 2007). Other approaches to extreme quantile estimation could and should be considered as well—as soon as they are implemented, they will fit our directional quantile scheme in the same vein. Although the block

maxima approach is also implemented in the R package `evir`, as an example we pick the function `gpd`, using a peaks over threshold approach to fit a generalized Pareto distribution. The extreme quantile envelopes using the data from Nepali children are shown in Figure 5.1.

The estimated extreme quantiles, for $p = 10^{-6}$, 10^{-5}, 10^{-4}, and 2×10^{-4} are confronted with the convex hull of the data, the empirical estimate for any $p \leq (2.33) \times 10^{-4}$. The plot seems to provide some information about the extent of extremality of the points labelled by 3110 and 4238; a closer inspection reveals that 3110 lies on the (2×10^{-3})-th directional quantile envelope, while 4238 on the (10^{-6})-th one. The real worth of this information is closely related to the same question regarding estimated extreme quantiles in the univariate case; nevertheless, if the pertinent discussion concludes favorably for some univariate alternative, then our methodology provides its viable multivariate extension.

5.3 Multivariate quantile regression

5.3.1 Bivariate Growth Charts

Directional quantiles can be easily extended to multivariate quantile regression. To simplify the illustration of our proposal without losing the generality, we will focus on bivariate growth charts, a case of bivariate regression quantiles. The general form of a growth chart is a series of smoothed quantile curves, showing how selected quantiles for the measurement, say, height or weight, change when plotted against age. In Figure 5.2, the growth charts used logarithms of primary variables, weight and height versus age with quantile indices $\{.03, .05, .10, .25, .50, .75, .90, .95, .97\}$. Subject 3110 is lying on about the 95% quantile curve in the left panel; while her height is sitting on the 25% quantile curve in the right panel. Her weight and height would be described as

"normal" if individually checking, but considering her low height position and relative high weight position among its peers, she may be a little overweight. The problem is to access what extent subject 3110 is overweight, or in another words, what is her position comparing with her peers according to the information of both weight and height.

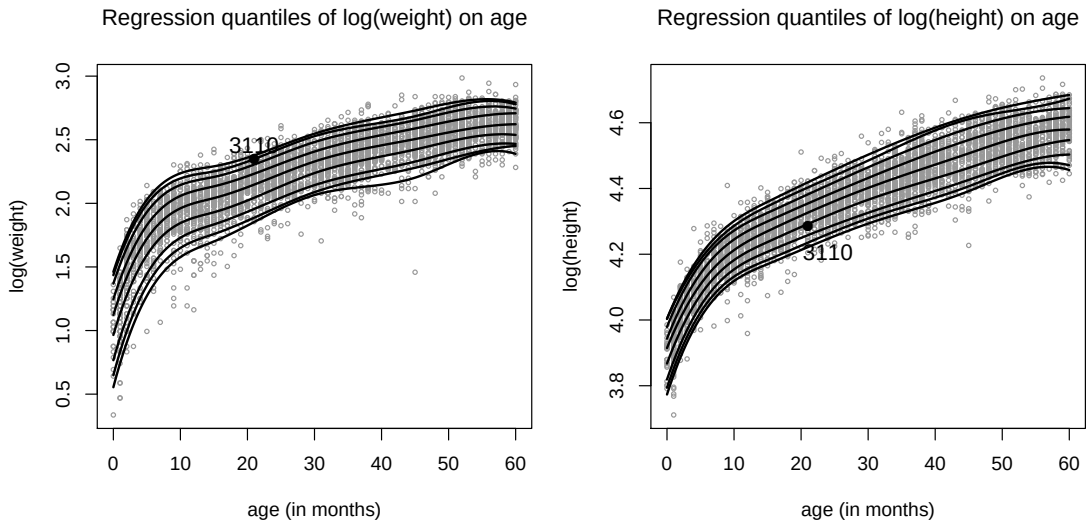

Figure 5.2: *The growth charts, quantiles of the primary variables, weight and height, regressed on the covariate, age.*

One way of dealing with this question is to look at the combination of height and weight, say Body Mass Index (BMI). In the left of Figure 5.3, it shows the quantiles of the logarithm of BMI (a projection of the vector response, the logarithms of weight and height) panel regressed on the covariate, which is the age in months. In left panel, subject 3110 is quite far away from its peers—lying above 97% quantile curves, in other words overweight, according to her BMI value higher than the values of her peers. An expert on nutrition may dispute the relevance of BMI for young children, and remind us of possible alternatives—for instance, the Rohrer index (ratio of weight to cubed height, hereafter ROI) and the weight-to-height ratio, the "growth charts" of ROI are displayed in the right panel of Figure 5.3. We do not think that the problem lies in deciding whether

that or another index is to be preferred; the essence of the data may lie well beyond the index-style of description. While such "growth charts" in Figure 5.3 are facilitating a lot of useful insights, we may like to confront them with a directional perspective—in a related covariate-dependent context.

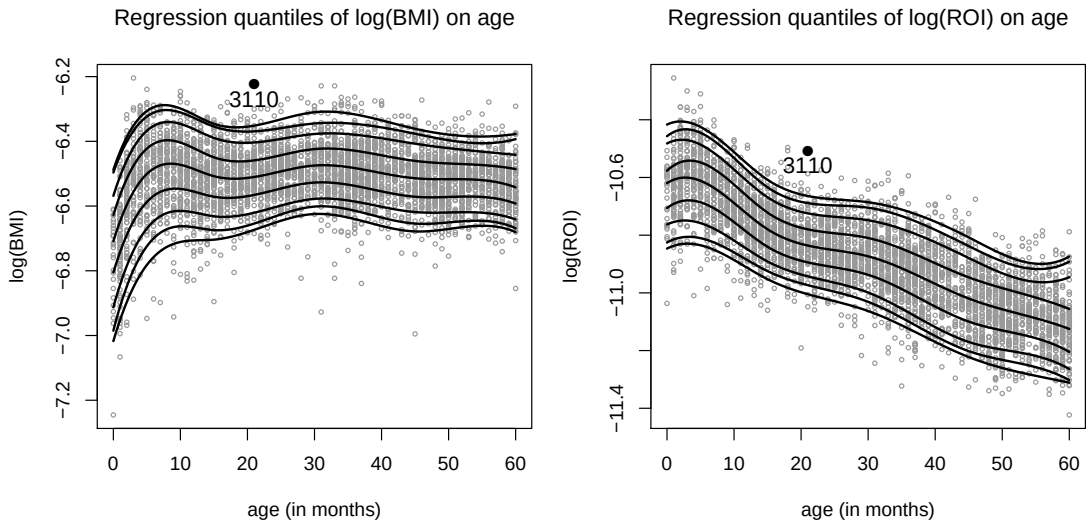

Figure 5.3: *The "growth charts", quantiles of logarithms of BMI and ROI, regressed on the covariate, age.*

Instead of exploring the individual index, we would rather study the joint information of weight and height. In particular, the concept of directional quantile envelopes we proposed provides the possibility of looking at the joint information of weight and height versus age—bivariate growth charts. The general form of a bivariate growth chart is a series of quantile tubes, corresponding to quantile envelopes for the bivariate measurements, changing with age. Plotting such an object stumbles upon the inevitable fact that our graphical universe is two-dimensional; animations and interactive graphics are certainly possible, but in traditional settings we can merely choose to plot directional quantiles for some fixed value(s) of the covariate—as in Figure 5.4, which shows the predicted envelopes for four values of the age (selected randomly except for age 21

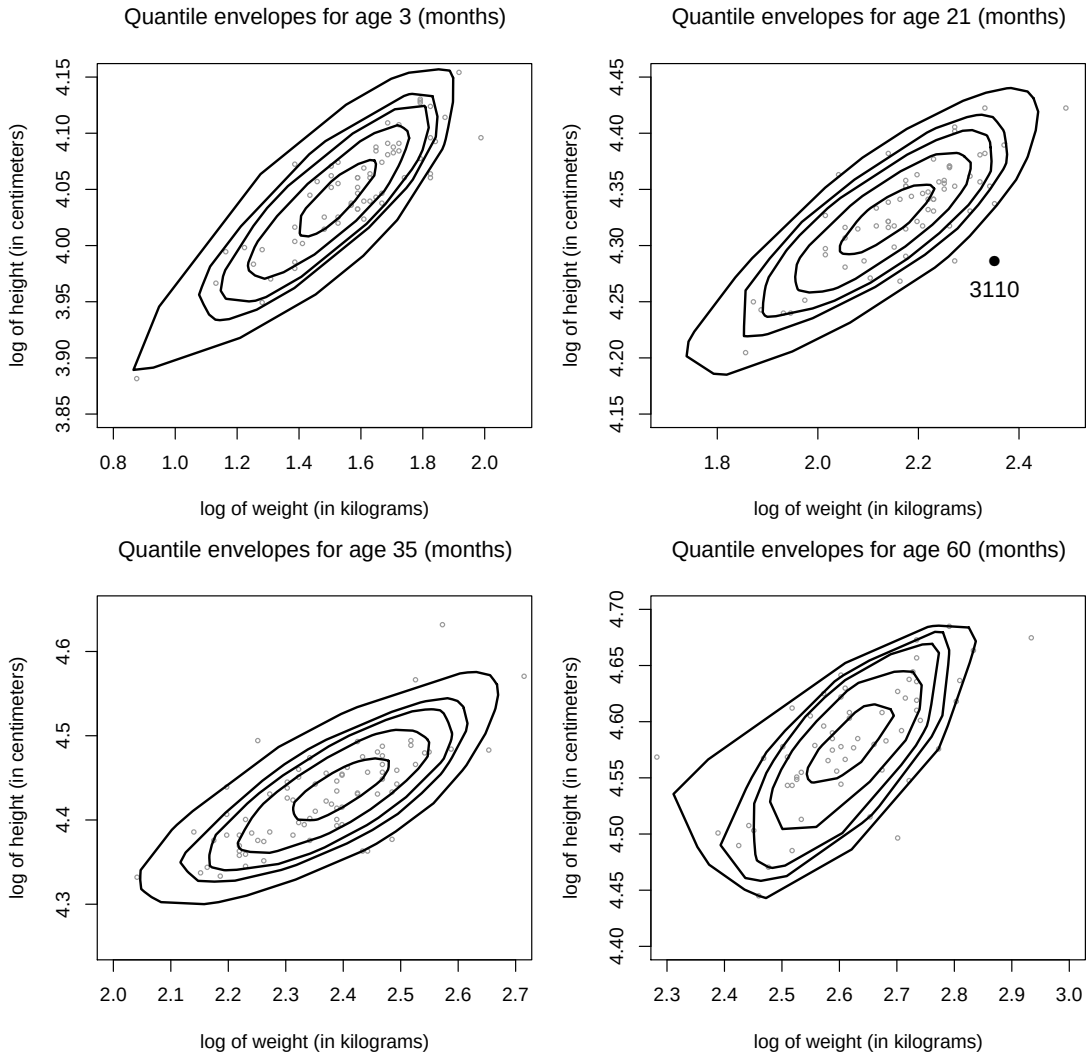

Figure 5.4: *Imagine the animation in which the directional quantile envelopes slowly ascend upward along the data cloud, demonstrating the dependence on the growing covariate, age.*

months, rather than pursuing any other objective). The four selected ages are 3, 21, 35 and 60 months with p values $\{.01, .03, .05, .10, .25\}$. The plotted datapoints in each panel represent the subjects with the particular age. In the upper right panel, subject 3110 is lying outside the $.01$th directional quantile envelopes at age 21 months. The interpretation of her position comparing to her peers can be obtained through directional

quantile lines with the specification of age 21 months. For instance, subject 3110 is either higher or lower (depends on selected directions) than 99% of her peers at age 21 months regarding any index which can be expressed by the linear combination of the logarithm of weight and height, which includes the indices of BMI and ROI.

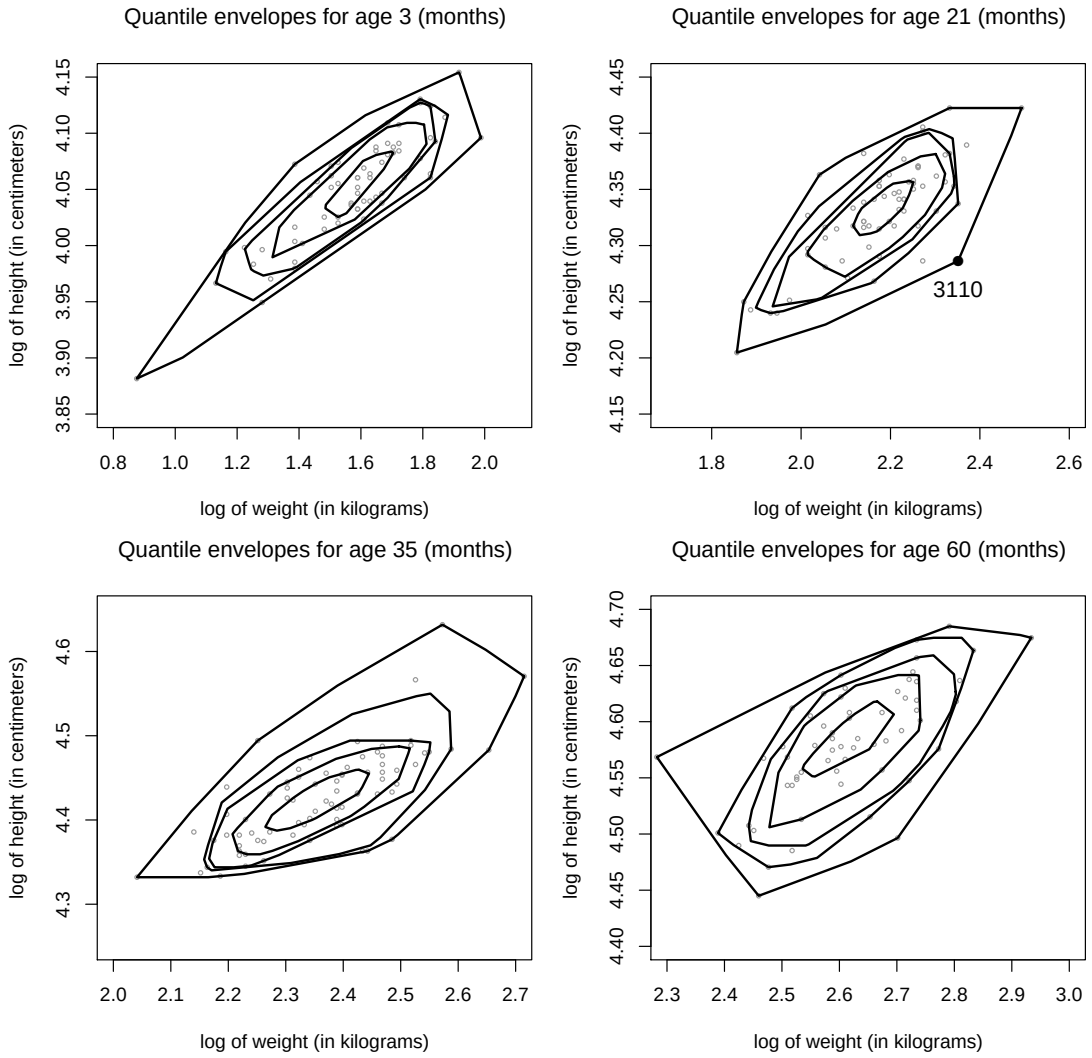

Figure 5.5: *If we computed directional quantile envelopes from these points at a particular age separately, the resulting contours would be rougher, and would vary from one value of age to another, which is against the nature of growth pattern.*

If we computed directional quantile envelopes from these points at a particular age separately, the resulting contours would be rougher, and would vary considerably from one value of age to another; see Figure 5.5. The quantile envelopes of weight and height versus age of a particular p value would not be a smooth tube, which is against the nature of the growth pattern of weight and height. The growth paths of weight or height is smooth as they change gradually according to time. The contours presented in Figure 5.4 borrow strength from other ages; we construct quantile envelopes using the information from all ages, similarly as we do in univariate regression.

The regression technique we used here is regression spline with specified smoothing parameter (smoothing degrees of freedom). The smoothing parameter was selected by eyeballing the plots, and then adopting a universal smoothing parameter for all directions. We are aware that while this may work well in certain situations (as it did in ours), one can easily imagine data exhibiting more signal-to-noise in one compared to other directions—then this fact should be reflected in variable smoothing parameters. For details, see the reviews in Section 1.2. Alternative nonparametric quantile regressions strategies are possible here; for some of those, see the reviews in Section 1.2.

As a summary, we will give a complete description of our proposal of multivariate quantile regression via directional approach using bivariate growth charts. The crucial part of our proposal is to project the bivariate response variable, say height and weight, into all directions and find the corresponding univariate quantile regression in those directions. In the univariate quantile regression, we use the projections as response variables and age as dependent variable, where various smooth methods can be applied. For any given age, we can calculate the predicted projection of height and weight through the obtained quantile regression line in all directions; then we can construct

the quantile envelope at that age. To obtain the bivariate growth charts we need to find the quantile envelopes at all ages for given p-values.

To fit univariate quantile regression, any method besides the ones described in Section 1.2 can be used, as long as it is suitable for univariate growth charts. The fitted bivariate growth charts are nested smooth tubes along age; for any particular age these tubes result in nested convex contours. The constructed bivariate growth charts also inherit good properties from univariate quantile regression. For example, if the univariate quantile regression includes an intercept, then the predictions of the response variable, projection of height and weight, are scale and translation equivariant; thus the estimated bivariate growth contours at any age are affine equivariant by Theorem 5.1.1. The most important thing is the estimated bivariate growth charts are able to provide more informative interpretation according to Theorem 4.3.1 than univariate growth charts.

5.3.2 Application: Finnish children longitudinal growth data

The data, see Pere (2000) for detailed description, were collected retrospectively from health centres and schools in Finland. They consist of longitudinal measurements on height and weight for 2514 children; supine length, rather than height, was measured for infants less than 2 years of age. The data have been edited to remove a small proportion of children with low or missing birth weight, twins or otherwise suspicious records. Growth charts were constructed only from the records with both height and weight specified. The re-evaluated data comprise 1143 boys and 1162 girls, all full term, healthy, singleton births with between 3 and 44 measurements per child, yielding 44169 height and weight records. Infants were measured roughly monthly before the age of 2, and annually or biannually thereafter. On average about 20 measurements of height and weight were made between the ages of 0 and 20.

The data were for two cohorts: one consisting of 525 boys and 571 girls born between

1954 and 1962 (94 percent between 1959 and 1961) followed until approximately age 19; the other of 618 boys and 591 girls born between 1968 and 1972 (99 percent between 1969 and 1971) followed until age 12.5. The children in the data covered more than 0.5 percent of Finns at the time period. The cohorts were fairly similar and therefore were pooled by sex.

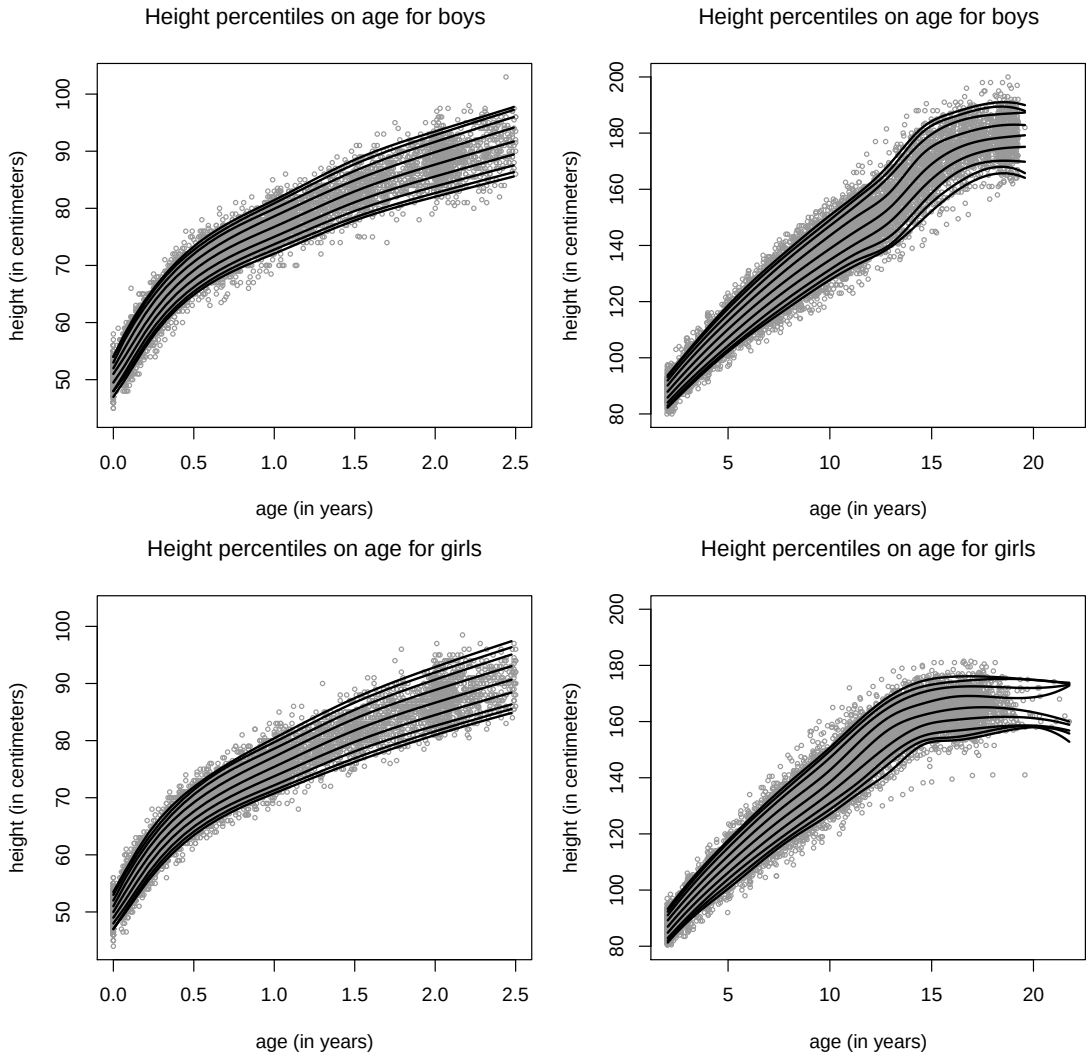

Figure 5.6: *Univariate growth charts of height for Finnish children using pre-selected internal knot sequence* $\{0.2, 0.5, 1.0, 1.5, 2.0, 5.0, 8.0, 10.0, 11.5, 13.0, 14.5, 16.0\}$.

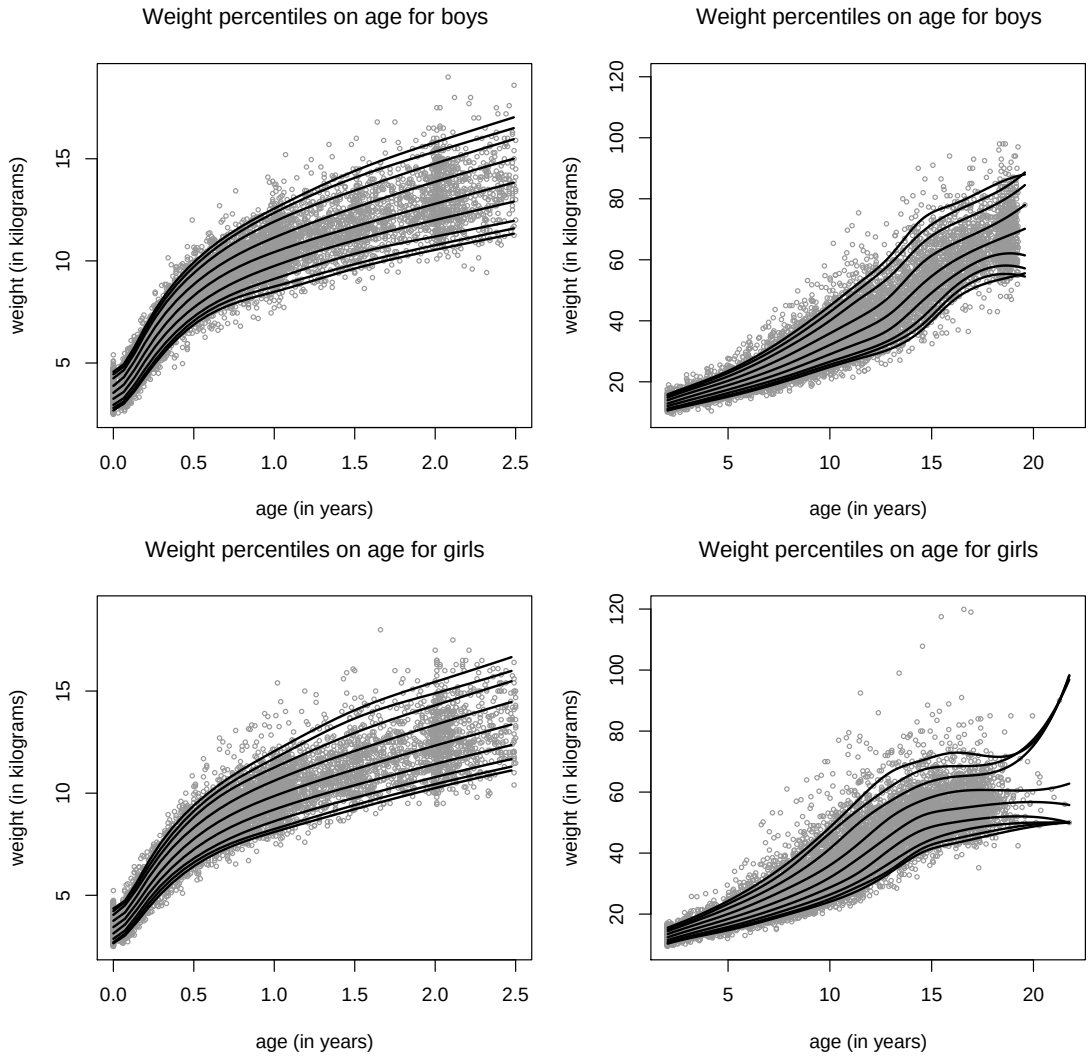

Figure 5.7: *Univariate growth charts of weight for Finnish children using pre-selected internal knot sequence* $\{0.2, 0.5, 1.0, 1.5, 2.0, 5.0, 8.0, 10.0, 11.5, 13.0, 14.5, 16.0\}$.

Similar to Wei, Pere, Koenker and He (2006), we accomplished our fits by regression splines the R package `splines` (R Development Core Team, 2008), and fitted quantile regressions by the R package `quantreg` (Koenker, 2007). The implementation of the quantile regression method employed the fixed set of cubic B-spline basis functions. To compare with the results of Wei, Pere, Koenker and He (2006), we used the original

scales of height and weight instead of taking the logarithm. We used the internal knot sequence $\{0.2, 0.5, 1.0, 1.5, 2.0, 5.0, 8.0, 10.0, 11.5, 13.0, 14.5, 16.0\}$, the same as Wei, Pere, Koenker and He (2006). Spacing of the internal knots is dictated by the need for more flexibility during infancy and in the pubertal growth spurt period. Linear combinations of these functions provide a simple and quite flexible model for the entire curve from birth to adulthood.

The univariate growth charts of height and weight are shown in Figure 5.6 and Figure 5.7. The growth charts of both height and weight distinguish boys from girls and infants from older children. The fitting around the boundary in both height and weight growth charts for boys and girls is not good. Some growth curves are crossing each other and some decrease around the boundary. As we have fewer observations around age 20 than other ages, it is hard to improve the quality of the growth charts.

Figure 5.8 displays the quantile contours at 1 and 12 years of age for girls and boys using pre-selected internal knots, overlaid with the subsample around each age points. Based on Figure 5.8, boys and girls have almost identical joint distribution of height and weight at the infant period; while at the puberty period, they have similar distribution of height and weight, but with different locations and scales, which confirms that boys and girls have a different growth path. We also can observe that the distribution of height and weight for boys and girls indeed changes over age in location, scale and shape.

To demonstrate the screening function of bivariate growth charts, we selected six illustrative subjects A-F from the Finnish children data, all boys at 1 year old of age. We plotted their height and weight against the estimated quantile envelopes at age 1 to locate their quantile ranks, as shown in Figure 5.9. Their height and weight remained outside $.03$-th quantile envelope except E between $.03$-th and $.05$-th quantile envelopes, indicating somewhat unusual body size. The directions of their outlyingness provide

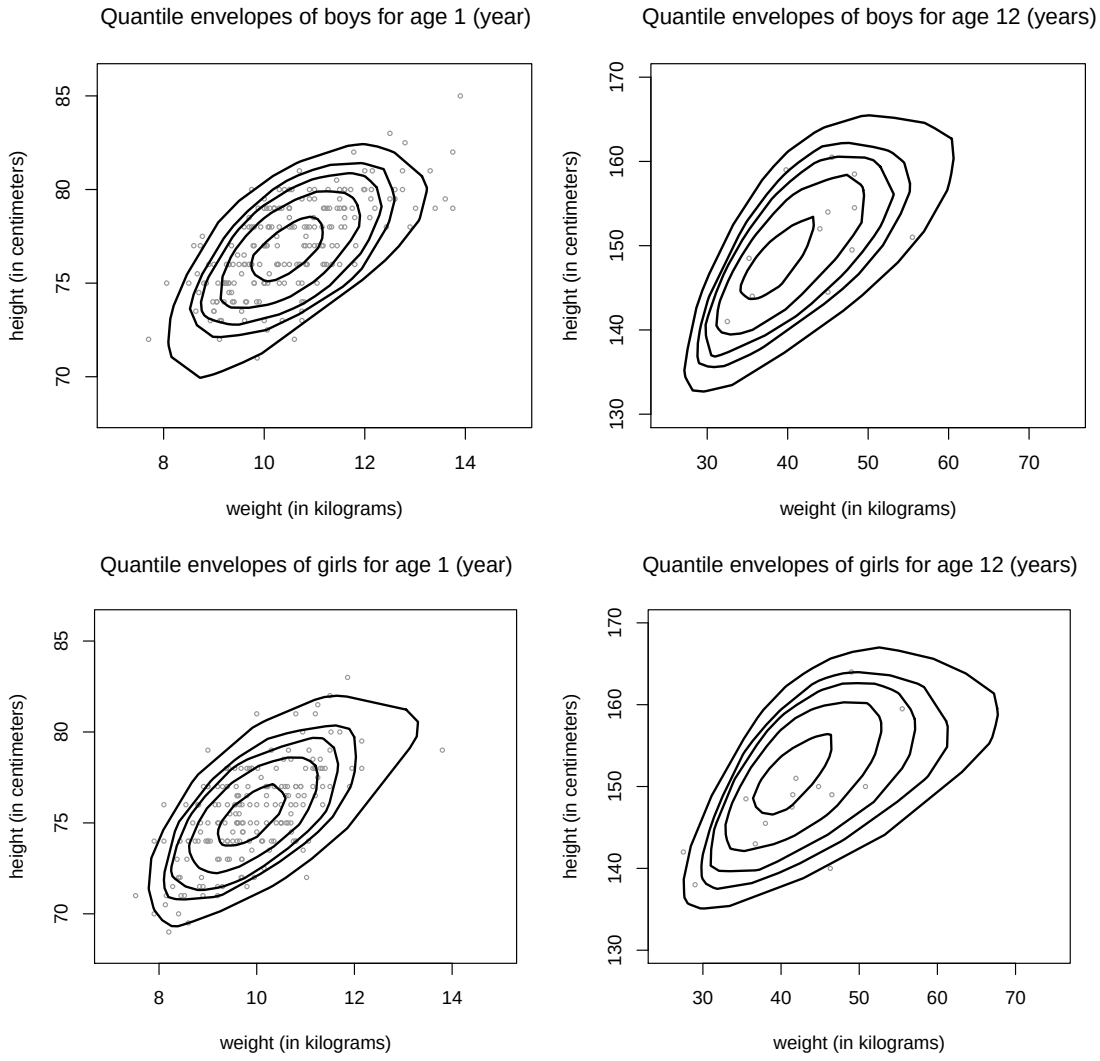

Figure 5.8: *Bivariate growth contours for boys and girls at* 1 *and* 12 *years old of age with* p *values* $\{.01, .03, .05, .10, .25\}$.

information on the way they had unusual boy size. Subject A had unusually small body size with both height and weight in the lower quantiles. In contrast, subject F had a relative larger body size than her peers; both her height and weight are greater than those of others. Subjects B, C, D, and E had normal height and weight, but subject B and C were underweight given their heights, while subject D and E were overweight given their heights.

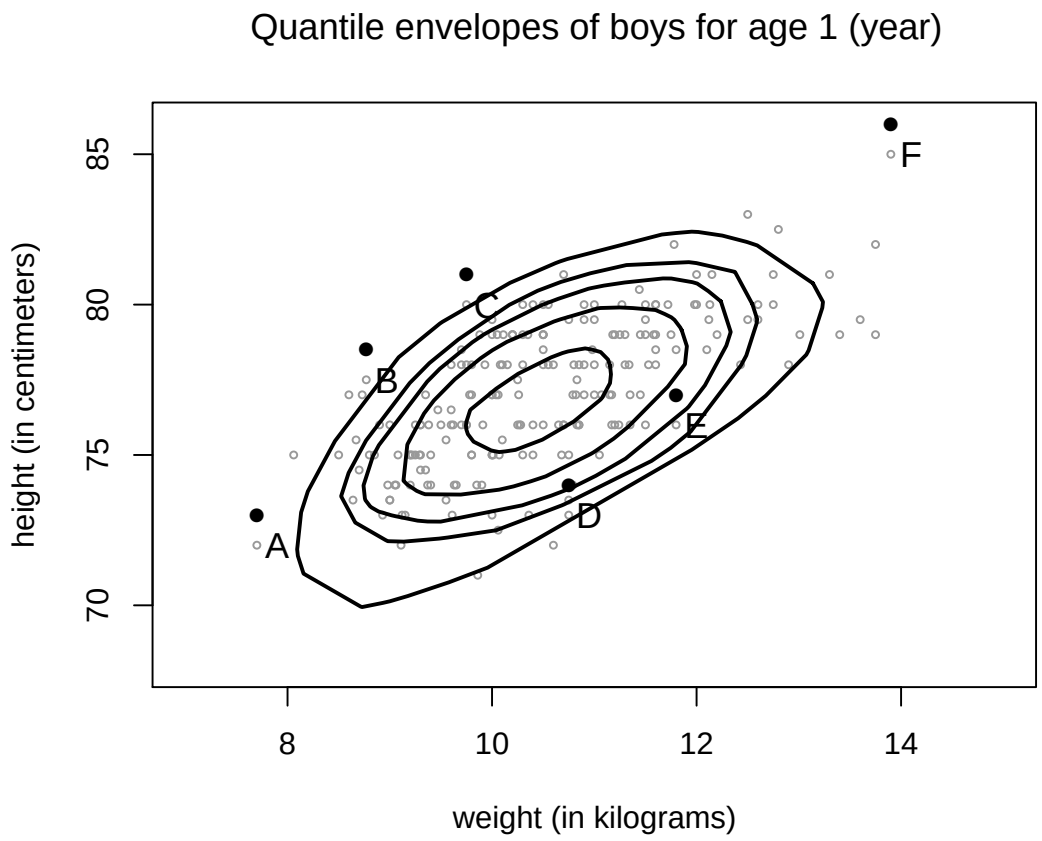

Figure 5.9: *Screening subjects A-F based on bivariate growth charts. From outside to inside, the quantile indices are* $\{.01, .03, .05, .10, .25\}$.

To compare with univariate growth charts, we also plotted height, weight, BMI growth charts and weight percentiles on height charts, superimposed with subjects A-F as displayed in Figure 5.10. Clearly, no univariate growth chart was as informative as the bivariate charts. For example, subject F was screened out in height and weight charts but not in BMI and weight on height charts. Subject B, C, D, and E were normal in height and weight charts, but were screened out in BMI charts and weight on height charts because of either their underweight or overweight features. The bivariate growth chart was able to effectively capture the type of abnormality of these subjects.

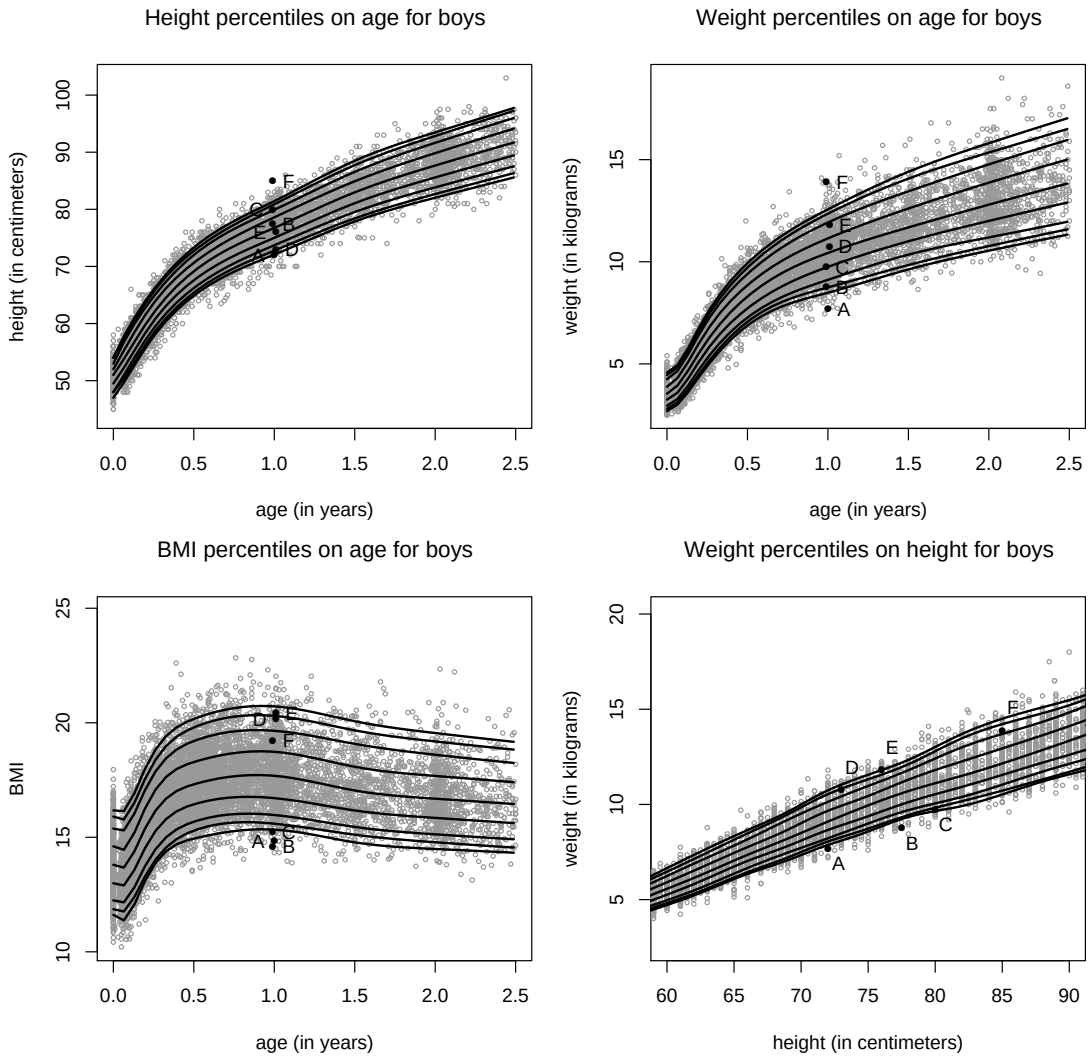

Figure 5.10: *Screening subjects A-F based on univariate growth charts. In each panel, the quantile indices are* $\{.03, .05, .10, .25, .50, .75, .90, .95, .97\}$.

5.3.3 Comparison of two methods for bivariate growth charts

As a special case of multivariate quantile regression, a typical bivariate growth chart has height and weight as the response variable and age as one of the covariates. Because of the request of meaningful interpretation of bivariate growth charts, most of the

multivariate quantile regression techniques reviewed in Section 2.4 are unsuitable for constructing the charts. The method of Wei (2008) stands up as a successful technique for bivariate growth charts. In this subsection we will compare the method of directional quantile envelopes with Wei's method.

There are two ways to index quantile contours as discussed in Section 4.1: indexing by the tangent mass or by the enclosed mass. Our directional quantile envelopes are indexed by the tangent mass. One advantage is the fact that the quantile envelopes indexed by the tangent very naturally interacts with marginal (and projected) quantiles, as shown in Figure 4.3. In the proposal of Wei (2008), quantile contours are indexed by the enclosed mass, or the "coverage". This indexing style also provides informative interpretation, but does not interact well with marginal (and projected) quantiles. It is still possible to construct directional quantile envelopes with prescribed coverage by employing a simple search. In fact, this approach was adopted by Rousseeuw, Ruts and Tukey (1999) for their bivariate boxplot, which is also feasible even in more sophisticated situations—for instance, in the quantile regression context.

Despite these possibilities, we do not advocate this way of indexing, but rather once again stress the strong interpretational appeal of indexing by the tangent mass. Even when realizing that the desideratum of "coverage" is codified in the daily standards of certain disciplines, we would still rather appeal to common sense—whether certain dogma cannot be changed. From this perspective, we view the approach of Wei (2008) rather complementary than alternative to ours. In what follows, we will use a simulation study and a real data study to compare the performance of Wei's method and our proposed method.

Suppose that $y_t = (y_{t1}, y_{t2})$ in $\mathbb{R}^2$ is a pair of measurements of interest at time t and that x is q-dimensional associated covariates. We assume that at any given time t and

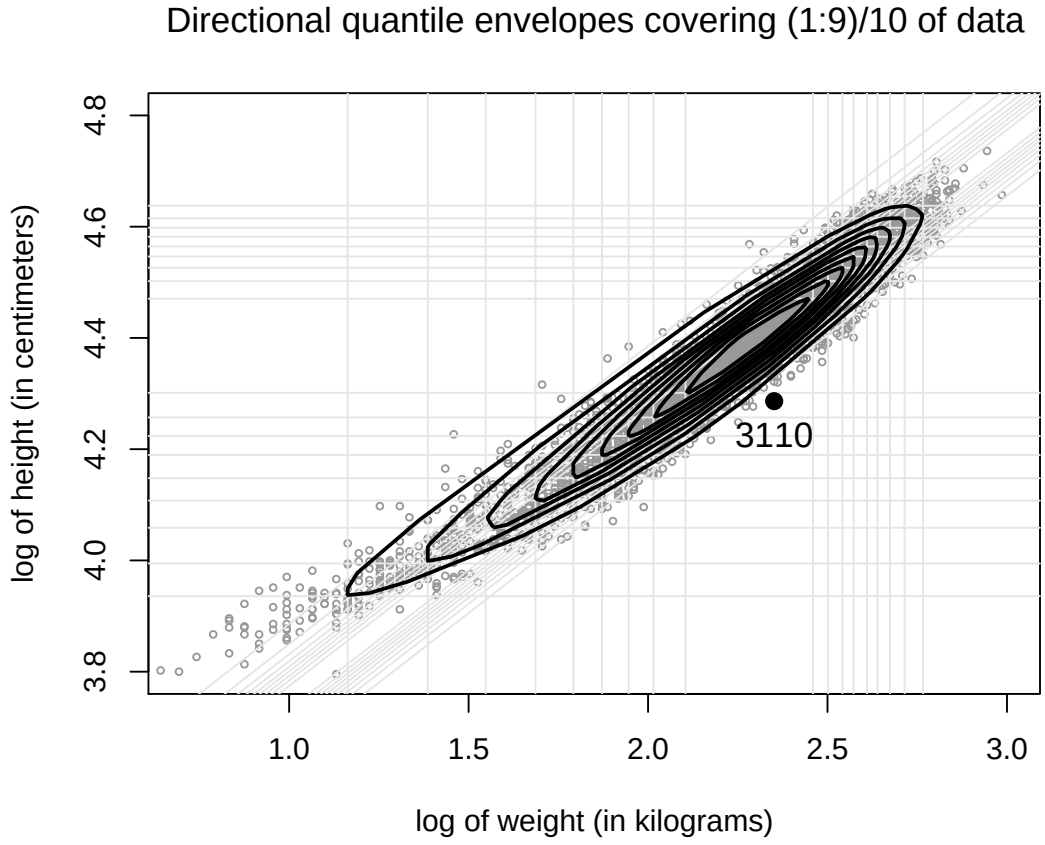

Figure 5.11: *It is still possible to find directional quantile envelopes with desired coverage, despite of the different underlying philosophy.*

covariate x, the measurement y_t follows a continuous joint distribution $F(y|t,x)$ on $\mathbb{R}^2$. We model $F(y|t,x)$ by constructing two stratified quantile models: a marginal model of y_{t1} and a conditional model of y_{t2} given y_{t1}; that is, for any $p \in (0,1)$, we specify

$$\begin{aligned} Q_p(y_{t1}) &= g_1(p;t,x), \\ Q_p(y_{t2}) &= g_2(p;t,x,y_{t1}), \end{aligned} \tag{5.3}$$

where we assume that $g_1(p;t,x)$ is the p-th quantile function of y_{t1} given time t and

covariate x, and $g_2(p; t, x, y_{t1})$ is the p-th conditional qauntile function of y_{t2} at time t and covariate x given the value of y_{t1}. The simulation theory we used is that if $(Y_1, Y_2) \sim F_{Y_1,Y_2}$, then $(F^{-1}_{Y_1}(U_1), F^{-1}_{Y_2|Y_1}(U_2|U_1)) \sim F_{Y_1,Y_2}$, where U_1 and U_2 are two independent random variables uniformly distributed on $(0, 1)$, F_{Y_1} is the marginal distribution of Y_1 and $F_{Y_2|Y_1}$ is the conditional distribution of Y_2 given Y_1. Therefore, we can use Rosenblatt's transformation (Mardia, Kent and Bibby, 1979) to simulate the pair (Y_1, Y_2) that follows the distribution of y_t given time t, and covariate x. Specifically, we draw U_1 from $(0, 1)$, and generate Y_1 from the first equation in 5.5 as the U_1-th quantile given t and x; then we draw U_2 from $(0, 1)$ and use the second equation in 5.5 to generate Y_2 as the U_2-th quantile at the same value of t, x and Y_1 obtained in the first step; for details, see Wei (2008).

In the simulation study, we consider two models, as did Wei (2008):

Model S1:

$$\begin{aligned} Y_1 &= 40t/(1+2t) + e_1, \quad e_1 \sim N(0,1); \\ Y_2 &= \ln(1+t) + (1+.2t)Y_1 + e_2, \quad e_2 \sim N(0,1) \end{aligned} \tag{5.4}$$

Model S2:

$$\begin{aligned} Y_1 &= 40t/(1+2t) + (1+.2t)e_1, \quad e_1 \sim SN(0,1,5); \\ Y_2 &= \ln(1+t) + (1+.2t)Y_1 + (1+.1Y_1)e_2, \quad e_2 \sim N(0,1) \end{aligned} \tag{5.5}$$

where SN is the skew-normal distribution (Azzalini and Capitanio, 2003). In both models the marginal mean curves of Y_1 and Y_2 are $40t/(1+2t)$ and $\ln(1+t) + 40t(1+.2t)/(1+2t)$, which resemble the paths of human growth trajectories on the log scale; for more discussion, see Wei (2008). We generated a Monte Carlo sample of size $5,000$ from model S1 and S2 with time t on the interval $(0, 3)$. In the first step of simulation, we fit the entire data with models

$$\begin{aligned} Q_p(y_1(t)) &= a_p(t), \\ Q_p(y_2(t)) &= b_p(t) + c_p(t)y_1(t), \end{aligned} \tag{5.6}$$

where $a_p(t)$, $b_p(t)$, and $c_p(t)$ are estimated by cubic regression splines with evenly spaced internal knots $(.5, 1, 1.5, 2.0, 2.5)$. The goodness of fit at various time points was accessed in Wei (2008), and the results suggest a good match between the empirical and model-based distributions. In the second step, we generated a sample of size $5,000$ from the estimated models 5.6 at each time of $t = .5, 1, 2$. We then reversed the order of Y_1 and Y_2 in 5.6 and repeated the two steps to obtain another sample of size $5,000$ at each time point. In the end, we combined the two samples at each time point to have a sample of size $10,000$.

With the simulated sample of size $10,000$, we construct the $.5$-th, $.75$-th, and $.95$-th quantile contours using Wei's method at each time point. The R package we used is `cobs`, implemented for the constrained quadratic smoothing B-spline by Ng and Maechler (2006). Using the same samples, we also construct the $.01$-th, $.05$-th, and $.10$-th quantile envelopes via directional approach. The elimination algorithm we used to estimate the envelopes here will be discussed in the next chapter. The results are displayed in Figure 5.12 and Figure 5.13, overlaid with their corresponding theoretical ones (in dashed lines); all the left panels are quantile contours from Wei's method, and all the right panels are quantile envelopes via directional approach. We can see that both the estimated quantile contours and the estimated quantile envelopes agree with the true ones fairly well under both models at different time points and different quantile indices. We may evaluate the difference between the estimated contours (envelopes) and the corresponding true ones using difference criterions, for instance, Pompeiu-Hausdorff distance; however, this is beyond the scope of this thesis. Nevertheless, it is fair enough to draw the conclusion that in Figure 5.12 and Figure 5.13 the agreement of envelopes and their true ones via a directional approach is not worse than the agreement of quantile contours and the true ones by Wei's method.

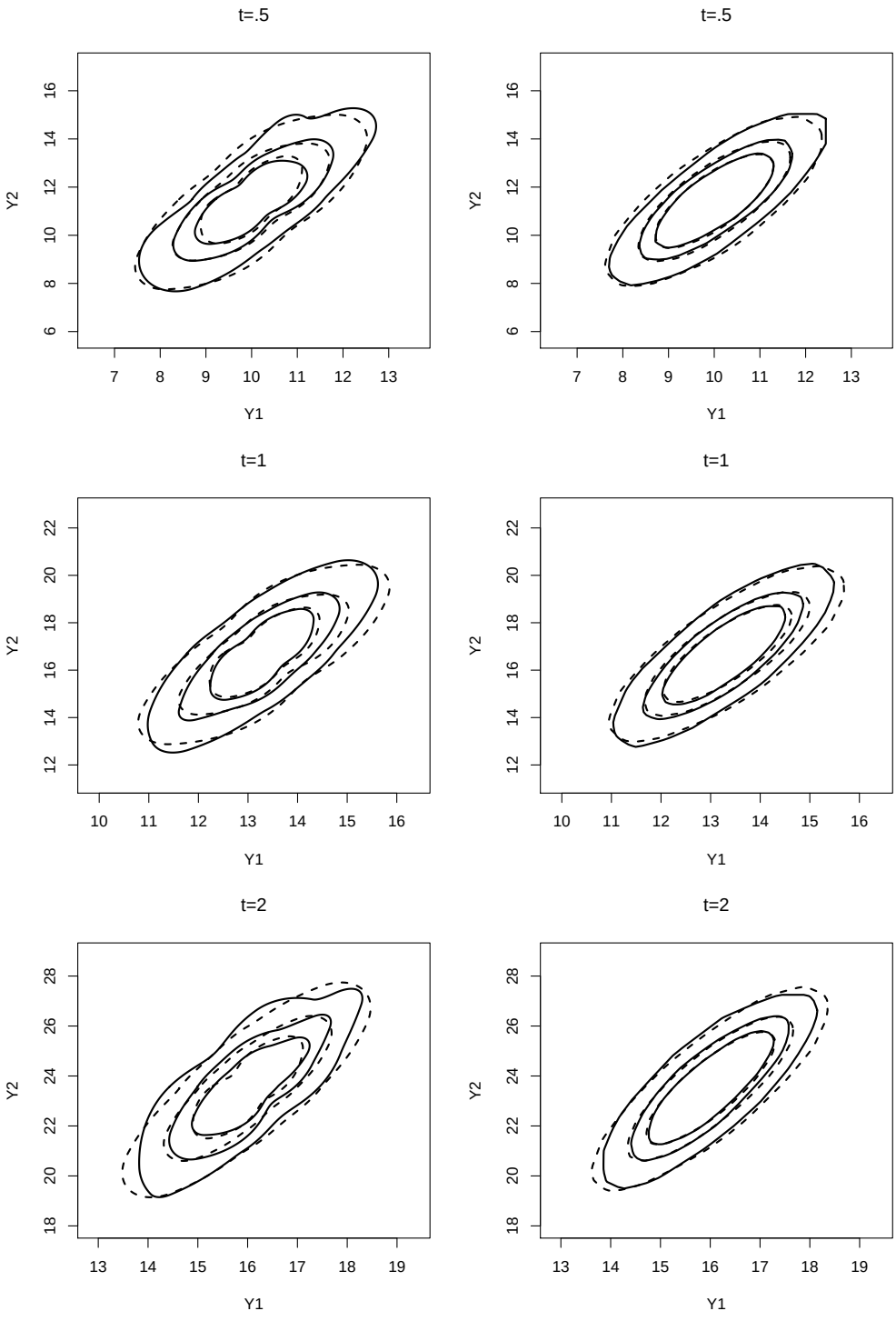

Figure 5.12: *Estimated quantile contours under model S1 from Wei's method with quantile indices* $\{.5, .75, .95\}$ *(left panels); and estimated quantile envelopes under model S1 via directional approach with quantile indices* $\{.01, .05, .10\}$ *(right panels). The dashed contours (envelopes) are the corresponding true ones.*

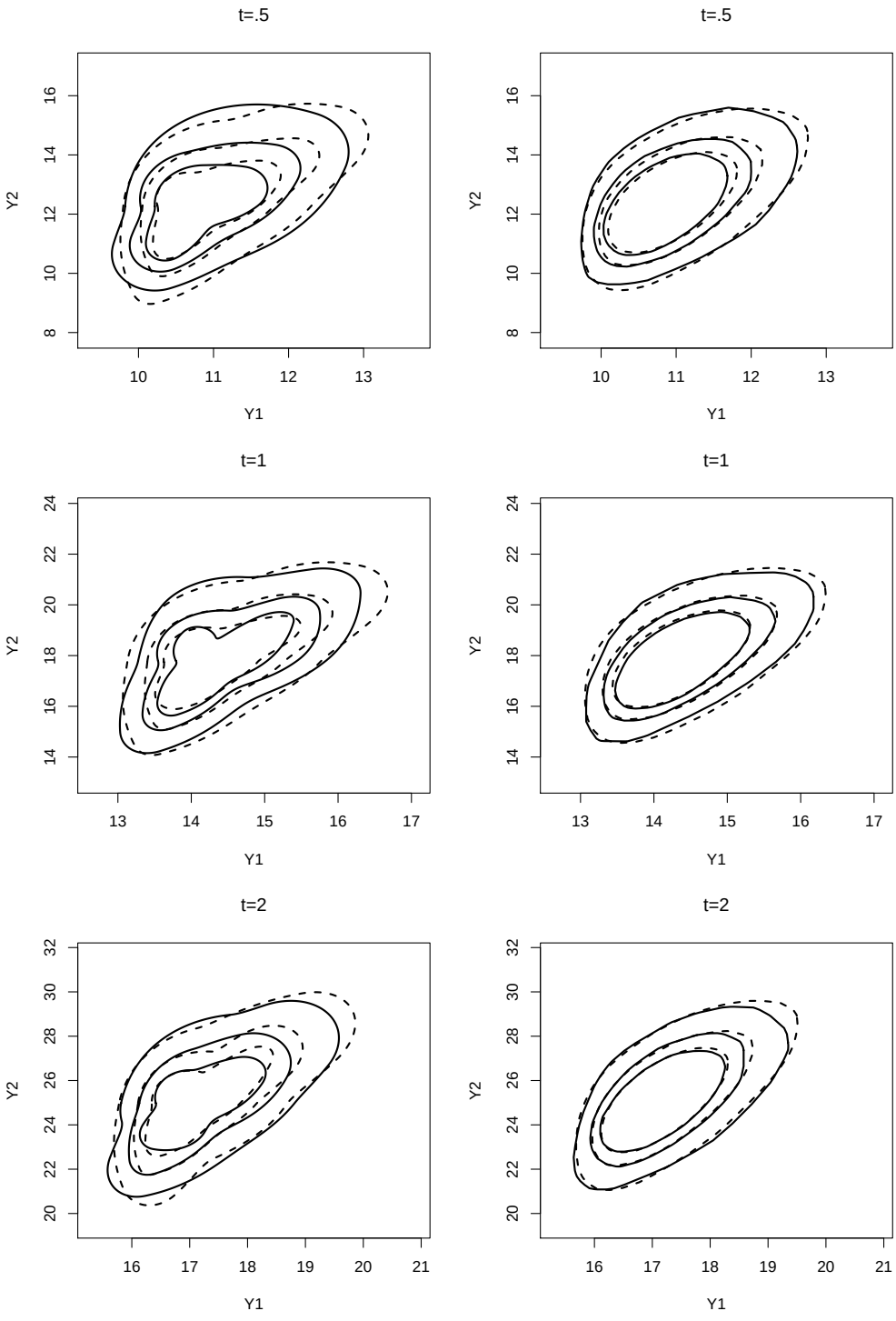

Figure 5.13: *Estimated quantile contours under model S2 from Wei's method with quantile indices* $\{.5, .75, .95\}$ *(left panels); and estimated quantile envelopes under model S2 via directional approach with quantile indices* $\{.01, .05, .10\}$ *(right panels). The dashed contours (envelopes) are the corresponding true ones.*

In the real data study, we used the Finish children longitudinal growth data to construct bivariate growth charts of height and weight. The simplest bivariate joint charts of height and weight depends only on time. Let (W_t, H_t) be a pair of weight and height at age t; we model them by

$$\begin{aligned} Q_p(\ln(H_t)) &= a_p(t), \\ Q_p(\ln(W_t)) &= b_p(t) + c_p(t)\ln(H_t). \end{aligned} \tag{5.7}$$

The goodness of fit of model 5.7 was assessed in Wei (2008) and indicated that model 5.7 is reasonable. We first used the Finish data to estimate the parameters $a_p(t)$, $b_p(t)$, and $c_p(t)$ in model 5.7. Then we simulated $10,000$ pairs of height and weight from the estimated model 5.7 at age 1 and 2 years. The estimated quantile contours under model 5.7 from Wei's method with quantile indices $\{.5, .8, .9\}$ (left panels) and the estimated quantile envelopes under model 5.7 via a directional approach with quantile indices $\{.01, .03, .05, .10, .25\}$ (right panels) at age 1 and 2 years, overlaid with nearby data points, are displayed in Figure 5.14. Compared to the plots in Figure 6 of Wei (2008), We obtained similar contours in the left panels of Figure 5.14 at age 1 and 2 years. The envelopes at age 1 year via the directional approach are almost identical to what was achieved in Figure 5.8. The envelopes at 2 years are similar to those at 1 year with slight changes in location, scale and shape.

Children's body growth is also affected by factors other than age, for example, parental height, which carry genetic information and have an influence in children's body size, which can not be ignored. It is important for bivariate growth charts to be capable of adding in other covariates in addition to age.We incorporate an individual subject's average parental height P into the joint height-weight growth charts by considering the following model:

$$\begin{aligned} Q_p(\ln(H_t)) &= a_p(t) + d_{1p}\ln(P), \\ Q_p(\ln(W_t)) &= b_p(t) + c_p(t)\ln(H_t) + d_{2p}\ln(P). \end{aligned} \tag{5.8}$$

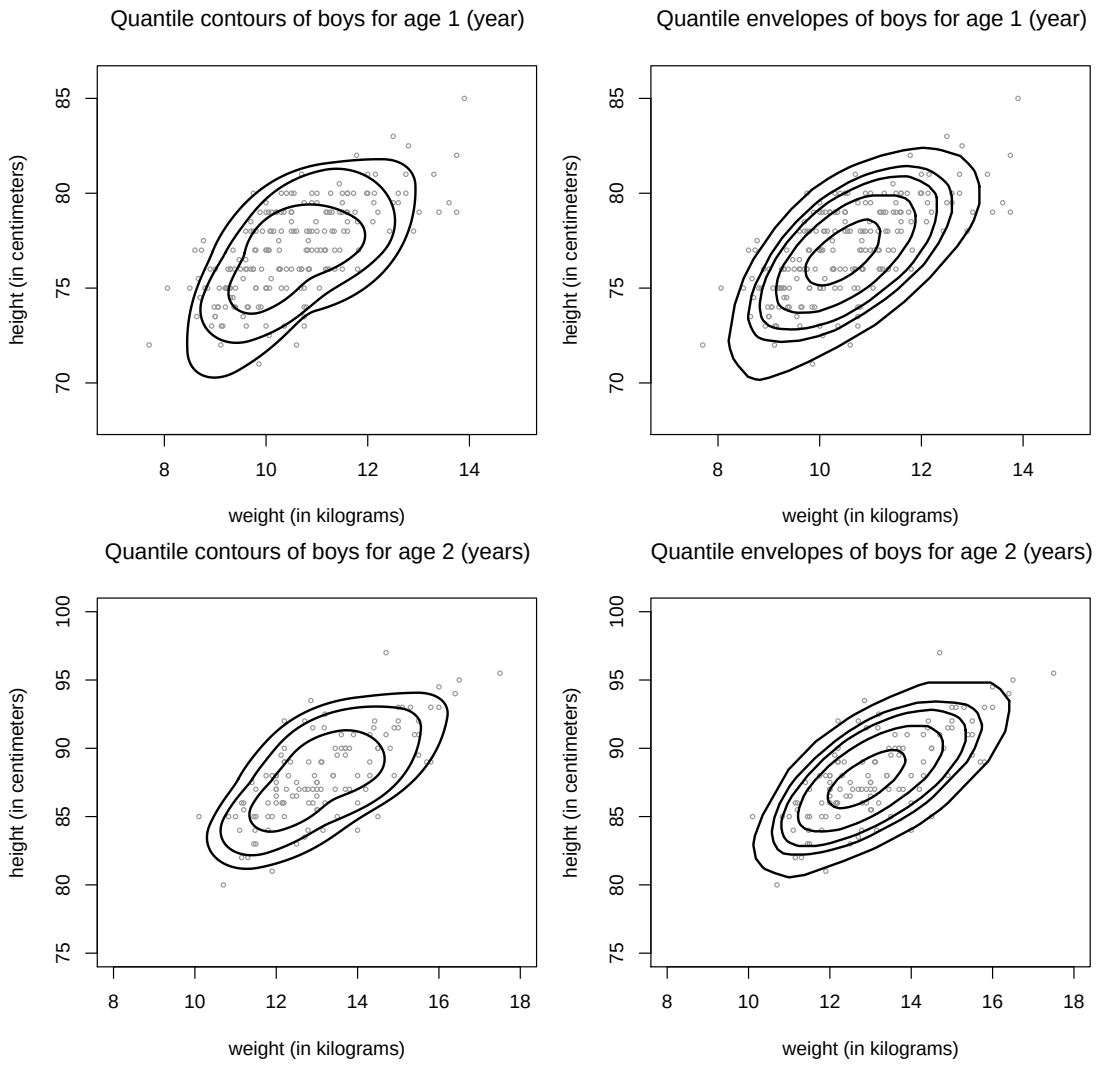

Figure 5.14: *Estimated quantile contours under model 5.7 from Wei's method with quantile indices* $\{.5, .8, .9\}$ *(left panels); and estimated quantile envelopes under model 5.7 via directional approach with quantile indices* $\{.01, .03, .05, .10, .25\}$ *(right panels).*

Following the same procedure, we estimated the parameters in 5.8, and generated a sample of size $10,000$ at age 2 years with average parental height 164 centimeters and 174 centimeters separately.

Figure 5.15 displays the estimated contours (envelopes) with and without considering parental heights. The upper panels are for the children with short midparent height, 165

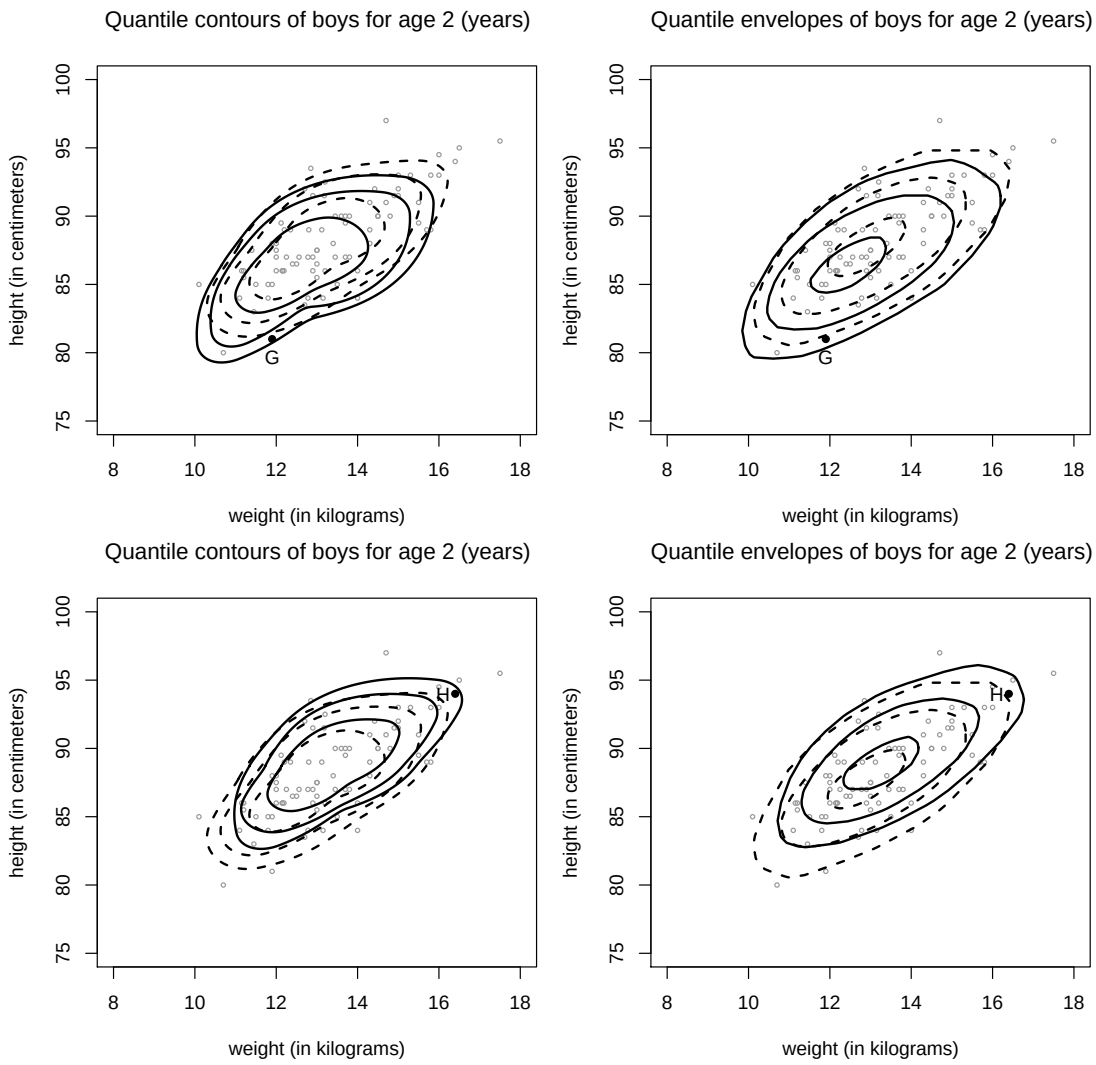

Figure 5.15: *Comparison of quantile contours (envelopes) with and without considering parental heights. Upper panels: midparent height is 165 cm. Lower panels: midparent height is 174 cm. The dashed contours (envelopes) are at age 2 years without considering parental heights, whereas the solid ones are their conditional version with inclusion of parental heights. The solid dot in the upper panels is the subject G with average parental height about 165 cm and the solid dot in the lower panels is the subject H with average parental height about 174 cm.*

centimeters and the lower panels are for the children with tall midparent height, 174 centimeters. The dashed contours (envelopes) are at age 2 years without considering parental heights, whereas the solid ones are their conditional version with inclusion of

parental heights. The quantile indices for contours by Wei's method from inside to outside are $\{.5, .8, .9\}$; the quantile indices for envelopes via the directional approach are $\{.10, .05, .01\}$. As shown in Figure 5.15, the quantile contours (envelopes) of the children with short parental heights shift toward the lower height and smaller weight end, which implies that those children generally have larger body size than the average of their peers. The shifts of both the quantile contours and envelopes of the children with tall parental height are toward the higher height and greater weight end. As Wei (2008) pointed out that the small body size of the children with short parents is reasonable and may be due to the "genes" inherited from parents; thus those children are much likely to be tall and heavy. Conversely, children with tall parents tend to have large body size. These conclusion can be drawn again from the quantile envelopes via the directional approach shown in Figure 5.15, which are originally suggested by the quantile contours in Wei (2008), similar to the contours in the left panels of 5.15. In a sense, our bivariate growth chart proposal via a directional approach and Wei's method are both flexible enough to add in covariates other than age.

Summarizing the comparison of our directional approach to bivariate growth charts and Wei's method, we can say the following. The two methods provide meaningful interpretations of the resulting envelopes (contours), which are complementary in scope: one by tangent mass and the other by enclosed mass. In the simulation study, both methods were capable of yielding envelopes (contours) which were a good approximation of the true ones. we demonstrated that our directional approach was flexible enough to add in covariates besides age, the way to obtain more informative readings than conventional univariate growth charts, BMI charts and Rohrer Index charts, as developed by Wei, Pere, Koenker and He (2005), Wei and He (2006), Wei (2008) (see also the discussion of Wei and He (2006)).

5.4 Approximation properties and consistency of estimators

There are several reasons to look at the effect of approximation on estimated quantile envelopes. The numerical motivation stems from the fact that in practice we do not take all directions to construct the directional quantile envelope; therefore, what is constructed is rather an approximate envelope $D_A(p)$ where A is a subset of selected directions, and we are interested in the quality of this approximation. We know that $D(p) \subseteq D_A(p)$, and we believe that a decent collection A of directions that reasonably fill $\mathbb{S}^{d-1}$ should make the approximation quite satisfactory. While the experimental evidence does not contradict this belief—for Figure 3.5 we used only 100 uniformly spaced directions, for Figure 3.6 we took 1009, and a negligible difference can be seen for $p = .10$—we would also like to have some theoretical support.

From the point of view of statistics, our directional quantiles are usually not the "true", but "estimated" ones. In the univariate case, it is not hard to conceive a standard general framework, in which under typical circumstances (say, independent sampling), the quantile estimates become more and more precise (say, with growing sample size), the latter formalized in some probabilistic sense. In this section, our concern is solely how such a consistent behavior of directional quantiles translates into an analogous behavior of directional quantile envelopes.

The following theorem, formulated in the general setting of supporting halfspaces, gives the essence of the relationship between directional quantiles and the envelopes constructed by them.

Theorem 5.4.1 *Suppose that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ is a sequence of closed sets with its union dense in a closed set $A \subseteq \mathbb{S}^{d-1}$, not contained in any closed halfspace whose boundary contains the origin. If for every sequence $s_n \in A_n$ that converges to $s \in A$,*

the sequence $q_n(s_n)$ converges to $q(s)$, then the sequence of sets $\bigcap_{s\in A_n} H(s, q_n(s))$ converges to $\bigcap_{s\in A} H(s, q(s))$ in the Pompeiu-Hausdorff distance—provided either the limit set is the closure of its interior, or it is a singleton and the sets in the sequence are nonempty.

Proof. To prove convergence with respect to the Pompeiu-Hausdorff distance, we exploit the following facts. First, the sequence $\bigcap_{s\in A_n} H(s, q_n(s))$, together with the limit $\bigcap_{s\in A} H(s, q(s))$ is contained in a bounded set, starting from some n. This follows from the fact that sets A_n are approaching a dense set in A, and the latter is not contained in any halfspace whose boundary contains the origin; therefore this property is shared by A_n starting from some n, which means that

$$\bigcap_{s\in A_n} H(s, \inf_{k\geq n} q_k(s))$$

is the desired bounded set. For uniformly bounded sequences, the convergence in the Pompeiu-Hausdorff distance follows from the convergence in the Painlevé-Kuratowski sense; see Rockafellar and Wets (1998), 4.13. The latter means that a general sequence of sets K_n converges to K if (i) every limit point of any sequence $x_n \in K_n$ lies in K; (ii) every point from K is a limit of a sequence $x_n \in K_n$. See also Mizera and Volauf (2002).

For sequences of closed sets with "solid" limits, sets that are closures of their interior, the Painlevé-Kuratowski convergence follows from the "rough" convergence, defined by Lucchetti, Salinetti and Wets (1994) to require (i) together with (ii)' every limit point of every sequence $y_n \in (\operatorname{int} K_n)^c$ is in $(\operatorname{int} K)^c$. See also Lucchetti, Torre and Wets (1993). That is, one can replace outer and inner convergence requirement of Painlevé-Kuratowski definition by two outer convergences, the original one, and the other one for "closed complements".

Suppose that $y \in \operatorname{int} K$. Then y belongs to all but finitely many K_n; otherwise, there would be a subsequence n_i such that $y \in (\operatorname{int} K_{n_i})^c$, and by the modified version of (ii)', $y \in (\operatorname{int} K)^c$. Hence, every y from the relative interior of K is a limit of an (eventually constant) sequence $y_n \in K_n$. To obtain (ii) for every $x \in K$, consider a sequence y_k of points from (nonempty) $\operatorname{rint} K$ such that $y_n \to y$; the desired sequence x_n is then obtained by a "diagonal selection": for every y_k, there is n_k such that $y_k \in K_i$ for every $i \geq k$; set $x_n = y_k$ for every $n_k \leq n < n_{k+1}$.

Thus, it is sufficient to prove (i) and (ii)'. Suppose that x is a limit point of a sequence $x_n \in \bigcap_{s \in A_n} H(s, q_n(s))$. Then there is a subsequence such that $s_n^T x_n \geq q_n(s_n)$ for every $s_n \in A_n$; every $s \in A$ is a limit of a sequence $s_n \in A_n$, therefore the assumptions of the theorem imply that $s^T x \geq q(s)$; hence $x \in \bigcap_{s \in A} H(s, q(s))$. This proves (i). This proves theorem for the singleton case, since then the Painlevé-Kuratowski convergence is implied by (i) once the sets in the sequence are nonempty.

Suppose now that x is a limit point of a sequence $x_n \in \left(\operatorname{int} \bigcap_{s \in A_n} H(s, q_n(s))\right)^c$, that is, a limit of some subsequence of x_n. Every such x_n satisfies $s_n^T x_n \leq q_n(s_n)$ for some $s_n \in A_n$. By the compactness of A, there is $s \in A$ that is a limit of a subsequence of s_n; passing to the limit along the appropriate subsequences, we obtain that $s^T x \leq q(s)$, by the assumptions of the theorem. This means that $x \in \left(\operatorname{int} \bigcap_{s \in A} H(s, q(s))\right)^c$. □

We can illustrate the use of this theorem on two instances. In the first, we take $q_n(s) = q(s) = Q(p, s)$; the theorem then says that the successive approximations, $D_{A_1}(p) \supseteq D_{A_2}(p) \supseteq \ldots$, approach $D_A(p)$ in the Pompeiu-Hausdorff distance. Typically, A_n are finite, while $A = \mathbb{S}^{d-1}$; the only requirements is that the directional quantiles $Q(p, s)$ depend on s in a continuous way—for instance, P satisfies the assumptions of Theorem 3.2.1. Basically, for a population distribution P satisfying certain conditions, by selecting a subset of A to obtain directional quantile envelopes

of the direction subset we can achieve an approximation of $D_A(P)$. The precision can be improved and even better the approximation arrives at the exact quantile envelope by adding more directions into the subset until it reaches A.

The second application furnishes a proof of consistency of $\hat{D}_n(p)$ to $D(p)$, when $\hat{D}_n(p)$ arise via applying the definition of directional quantile envelopes to empirical distributions P_n that converge weakly almost surely to the sampled population distribution P (under suitable sampling scheme like independent sampling). Since the consistency of depth contours was discussed more thoroughly by He and Wang (1997), we consider this rather an example of the use of Theorem 5.4.1. The required assumptions are those of Theorem 3.2.1, continuous or bounded support of P, and the nondegeneracy of the limit $D(p)$ (in general we cannot guarantee that $\hat{D}_n(p)$ are nonempty). The Skorokhod representation yields random variables X_n converging almost surely to random variables X, such that the laws of X_n and X are the corresponding empirical distributions P_n and P, respectively; Theorem 3.2.1 then implies the convergence assumption required by Theorem 5.4.1. More specially, we can pick $A_1 = A_2 = \cdots = A$, and if $s_n \to s$ holds true for the sequence s_n then $Q(p, s_n) \to Q(p, s)$ is guaranteed by Theorem 3.2.1. Consequently, $\hat{D}_n(p) \to D(p)$ in the Pompeiu-Hausdorff distance by Theorem 5.4.1.

To obtain some idea of the magnitude of the approximation error, we can proceed as follows. For simplicity, we limit our scope to the two-dimensional setting. Let D be a convex set and d a point on its boundary, i.e. $d \in \partial D$. The directions of all tangents of convex set D passing through point d generate a convex cone, $T_D(d)$. Let $c_D(d)$ be the maximal cosine of the angle between two directions in $T_D(d)$, then it is the cosine of the angle between two extremal normalized directions in $T_D(d)$, that is,

$$c_D(d) = \sup\left\{\frac{s^T t}{\|s\|\|t\|} : s, t \in T_D(d)\right\} = \sup\left\{s^T t : s, t \in T_D(d) \cap \mathbb{S}\right\}.$$

In fact, this cosine is the same as the maximal cosine of the directions in the normal cone $N_D(d)$, normalized version of $c_D(d)$; see Rockafellar and Wets (1998), Chapter 6. We can see that $c_D(d)$ is nonnegative and $c_D(d) \leq 1$, the equality holding if and only if $T_D(d)$ consists of a single direction, that is when D has a unique tangent at point d. Let

$$\kappa_D(d) = \sqrt{\frac{2}{1 + c_D(d)}},$$

and α be the angle between the two directions in $T_D(d)$ with maximal cosine. $\kappa_D(d)$ can be rewritten as $1/cos(\alpha/2)$, the reciprocal of the cosine of the half of the maximal angle between directions in the tangent cone $c_D(d)$. Denote $\kappa_D = \sup_{d \in \partial D} \kappa_D(d)$. Apparently, $\kappa_D \geq 1$, the equality holding true only for smooth D. On the other hand, κ_D can be equal to $+\infty$ for the degenerate D, the sets with empty interior.

Theorem 5.4.2 *Let $A \subseteq \mathbb{S}^1$ be a set of directions, and let $\hat{q}(s)$ and $q(s)$ be two functions on A. Suppose that both $\hat{D} = \bigcap_{s \in A} H(s, \hat{q}(s))$ and $D = \bigcap_{s \in A} H(s, q(s))$ are nondegenerate; then both $\kappa_{\hat{D}}$ and κ_D are finite and*

$$d\big(\hat{D}, D\big) \leq \max\{\kappa_{\hat{D}}, \kappa_D\} \sup_{s \in A} |\hat{q}(s) - q(s)|,$$

where d denotes the Pompeiu-Hausdorff distance.

Proof. As $\hat{D}$ and D are compact convex sets, we have $d(\hat{D}, D) = d(\partial\hat{D}, \partial D)$. Let $\epsilon = \sup_{s \in A} |\hat{q}(s) - q(s)|$; we will show that for any $x \in \partial D$, $d(x, \partial\hat{D}) \leq \kappa_D \epsilon$. Let $\tilde{q}(s) = q(s) - \epsilon$ and $\tilde{D} = \bigcap_{s \in A} H(s, \tilde{q}(s))$.

For simplicity, we assume that $\tilde{D}$ is also nondegenerate. We have that $\tilde{D} \subseteq \hat{D}$, and also $\tilde{D} \subseteq D$, the latter set being congruent to $\tilde{D}$. If $\kappa_D(x) > 1$, then x is a vertex of D. Since $d(x, \tilde{x}) = \kappa_D(x)\epsilon$, where $\tilde{x}$ is the corresponding congruent vertex in $\partial\tilde{D}$, it follows that $d(x, \partial\tilde{D}) \leq \kappa_D \epsilon$. When $\kappa_D(x) = 1$, then, by Theorem 24.1 of of Rockafellar (1996), there exists a sequence $x_n \neq x$, $x_n \in \partial D$, such that $x_n \to x$

and $s_n \to s$, $\kappa_D(x_n) = 1$, where s_n and s are the directions of the tangent lines passing through x_n and x, respectively. There are two possibilities.

If there is N such that $s_n = s$ for any $n > N$, then there must be two points, denoted by y_1 and y_2, in $\partial H(s, q(s)) \cap \partial D$ such that $\kappa_D(y_1) > 1$ and $\kappa_D(y_2) > 1$. That is, y_1 and y_2 are two vertices of D and there is no other vertex between y_1 and y_2 of D. Suppose that $\tilde{y}_1$ and $\tilde{y}_2$ are points congruent to them on $\tilde{D}$; then $\tilde{y}_1$ and $\tilde{y}_2$ are two vertices of $\tilde{D}$ and there is no other vertex between $\tilde{y}_1$ and $\tilde{y}_2$ of $\tilde{D}$ as well. In other words, we have a trapezoid with vertices y_1, y_2, $\tilde{y}_1$ and $\tilde{y}_2$ and x lies on one of the bases. A simple geometric calculation then shows the existence of a point, y, lying on the base constructed by $\tilde{y}_1$ and $\tilde{y}_2$, such that $d(x, y) \leq \max\{\kappa_D(y_1), \kappa_D(y_2)\}\epsilon$, that is, $d(x, \partial\tilde{D}) \leq \kappa_D\epsilon$.

Suppose that there is an infinite subsequence of s_n such that $s_n \neq s$. Let $\partial\tilde{D}$ by $\tilde{x}_n$ and $\tilde{x}$ be the congruent counterparts of x_n and x, respectively; let s_n and s be the corresponding directions. Let $y_n = \partial H(s_n, q(s_n)) \cap \partial H(s, q(s))$ and $\tilde{y}_n = \partial\tilde{H}(s_n, q(s_n)) \cap \partial\tilde{H}(s, q(s))$. We have that $y_n \to x$, $\tilde{y}_n \to \tilde{x}$, and $d(y_n, \tilde{y}_n) = \sqrt{2}\epsilon/\sqrt{1 + s_n^T s}$. As $d(y_n, \tilde{y}_n) \to d(x, \tilde{x})$ and $\sqrt{2}\epsilon/\sqrt{1 + s_n^T s} \to \epsilon$, we arrive to $d(x, \tilde{x}) = \epsilon$, which means $d(x, \partial\tilde{D}) \leq \kappa_D\epsilon$ again.

Taking into account that $\tilde{D} \subseteq \hat{D}$, we obtain that $d(x, \partial\hat{D}) \leq \kappa_D\epsilon$, for any $x \in \partial D$. The theorem follows from this and the symmetric inequality, $d(x, \partial D) \leq \kappa_{\hat{D}}\epsilon$ holding true for any $x \in \partial\hat{D}$, which can be established in an analogous way. □

Chapter 6

Computation issues [1]

6.1 Envelopes of directional quantile lines

Recall Definition 3.3.1; the p-th directional quantile envelope of a probability P is $D(p) = \bigcap_{s\in\mathbb{S}^{d-1}} H(s, Q(p,s))$, where $H(s,q) = \{x : \ s^T x \geq q\}$ is the supporting halfspace determined by a direction $s \in \mathbb{S}^{d-1}$ and p-th quantile $q \in \mathbb{R}^d$ on that direction. That is, the envelope of the directional quantiles is the intersection of some supporting halfspaces. From Section 5.4, we can select some directions and use the envelope of the quantile lines on the selected directions to approach the true envelope. The approximation can be made more accurate by the addition of more directions. Obtaining the envelope of directional quantiles is equivalent to finding the intersection of halfspaces whose boundary is directional quanile lines. In practice, we may just need an approximate envelope, which is equivalent to calculating the intersection of finite halfspaces.

In what follows we will focus on the calculation of the intersection of halfspaces in the plane. Given a collection of finite halfspaces in the plane, the intersection of them, if not empty and bounded, is a convex polygon. A convex polygon is uniquely determined by all its vertices, or extremal points. An intuitive idea is to find all the

[1] Chapter 6 with an R package, co-authored with Dr. Ivan Mizera, will be submitted for publication.

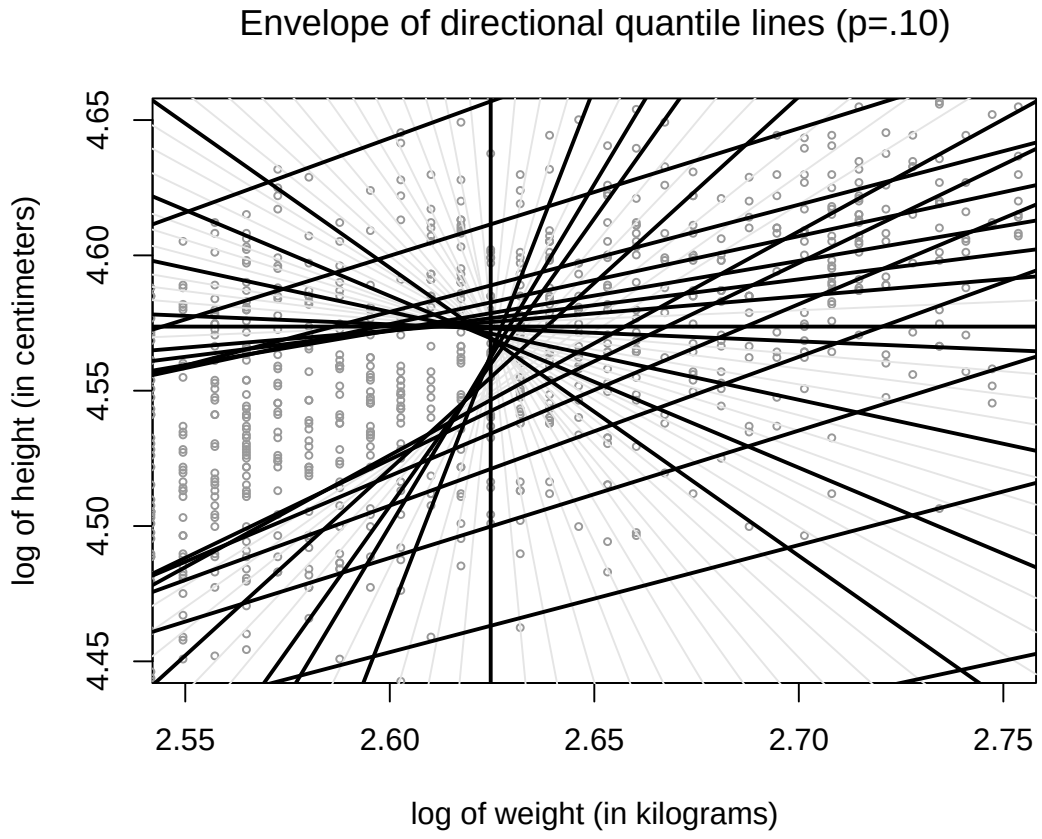

Figure 6.1: *Amplification of the left upper corner of the envelope in Figure 3.5. Some quantile lines, all grey lines, do not actively contribute to the envelope.*

intersections of the boundaries of halfspaces, that is, all the vertices. In summary, we are facing an obvious, nevertheless some effort requiring task: given a collection of finite halfspaces in the plane, construct a vertex representation of a convex polygon that is the intersection of these halfspaces. Such a representation consists of the collection of the extremal points of the polygon, circularly ordered so that each two in a sequence determine a face of the polygon. To implement this task, one crucial step is to find and remove the boundary lines of halfspaces which do not actively contribute to the polygon, all grey quantile lines as shown in Figure 6.1. We call this step a *validation procedure*;

it will be discussed in detail in the next section. The rest is relatively simple, calculating the intersection points of all boundaries, which are actually the extremal points of the polygon, and connecting them in order.

6.2 The elimination algorithm

Halfspaces are described by homogeneous coordinates with three components. The first two components, A and B, correspond to a nonzero vector $[A, B]$ perpendicular to the boundary of the halfspace and pointing inside it. The third component, C, is selected so that

$$Ax + By - C = 0 \tag{6.1}$$

is an equation of the boundary line of the halfspace. We write the homogeneous coordinates in the form $[A : B : C]$, the colons to indicate that $[kA : kB : kC]$ represents the same halfspace for every $k > 0$; for $k < 0$, the equation (6.1) still represents the same line but the halfspace is the opposite one.

If the vector $[A, B]$ has unit length, then C is the signed distance, in the sense of $[A, B]$, of the line (6.1) to the origin. Homogeneous coordinates are particularly convenient for computing intersections; the intersection of the boundary lines of halfspace $[A_1; B_1, C_1]$ and $[A_2; B_2, C_2]$ is obtained as the solution of the linear system

$$\begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}. \tag{6.2}$$

Like we mentioned earlier, the intersection of a collection of finite halfspaces is a convex polygon, if not empty and bounded. Each face of the polygon is a subset of some *edge*, which is how we name the oriented boundary lines of the halfspaces; the orientation of the edge of the halfspace $[A : B : C]$ is chosen by the "right-hand rule" to be that of $[B, -A]$, the vector $[A, B]$ rotated clockwise (if we place the right hand with

the palm down so that the extended thumb points into the halfspace, then the fingers show the orientation of the edge). The construction of the intersection of the halfspaces amounts to finding out which edges actively determine the polygon.

The algorithm takes as its input a sequence of homogeneous coordinates of halfspaces, assuming that the directional vectors are pairwise distinct and ordered in the counter-clockwise sense. They may contain pairs that are collinear, but then such pairs must have opposite directions. It is also assumed that the intersection of the halfspaces is bounded, which means that no two neighboring directional vectors may have the angle π or more in the counter-clockwise sense. The algorithm starts by taking the first edge, setting it as the current, `ecur`, and also marking it as the terminal, `etrm`, edge. The algorithm then proceeds in the counter-clockwise order on the bidirectional circular list of edges. Its basic step is the validation of the current edge—a decision which determines whether the edge is going to be active in the determination of the constructed polygon. The validation is based on the order of the outgoing and incoming vertices, `vprec` and `vnext`, on the current edge. If the edge is validated, the algorithm proceeds by declaring its successor the current edge. In this case the algorithm is moving forward, or in the counter-clockwise order. If the edge is not validated, then the algorithm moves backward: the current edge is eliminated, the previous edge is declared current, and the circular list is updated. In this case, the algorithm is moving in the clockwise order. The algorithm keeps the track of the number `nact` of active edges; if an edge is eliminated, this number is decreased by one.

Every vertex is an intersection of two edges, an *incoming* and *outgoing* one—in agreement with the orientation on the edges. For instance, the outgoing vertex `vprec` is the intersection of the outgoing edge, the preceding edge, and the incoming edge, the current edge; the incoming vertex `vnext` is the intersection of the outgoing edge,

now the current edge, and the incoming edge, now the next edge. The correspondence between potential vertices and edges is chosen so that every vertex is linked to its incoming edge. In this sense, the vertex `vprec` is linked to the current edge and the vertex `vnext` is linked to the next edge.

To validate the current edge, we need its outgoing and incoming vertices, `vprec` and `vnext`. Typically, `vnext` has to be computed, while `vprec` is the `vnext`for some previous step, and is already computed. The exceptions are: the terminal state (when the terminal edge is again declared current, and toggle `invl` is turned off), then the intersections are already computed; and the initial stage, when also `vprec` needs to be computed—a situation that may also arise in the backward motion when the former initial edge is invalidated.

The line intersection routine we use here is adapted to our special situation. We first need to check the angle, in the counter-clockwise sense, between the edges. If this is found not less than π, then the routine sends a signal to the algorithm to terminate by declaring the polygon empty. Otherwise, the intersection is computed; the angle condition guarantees that the system (6.2) is non-singular.

The validation is based on the following geometrical principle: if the intersections (`vprec` and `vnext`) are not equal, and the direction from `vprec` to `vnext` is opposite to the direction of the edge, then the halfspace is not active and toggle `invl` is turned on, see the left panel of Figure 6.2; on the other hand, if the direction from `vprec` to `vnext` is consistent to the direction of the edge, then the halfspace is active and toggle `invl` is not turned on, see the right panel of Figure 6.2. In the case when the intersections (`vprec` and `vnext`) are equal, the halfspace is not active either and toggle `invl` is turned on as well.

After the validation step, the algorithm either proceeds to the succeeding edge—

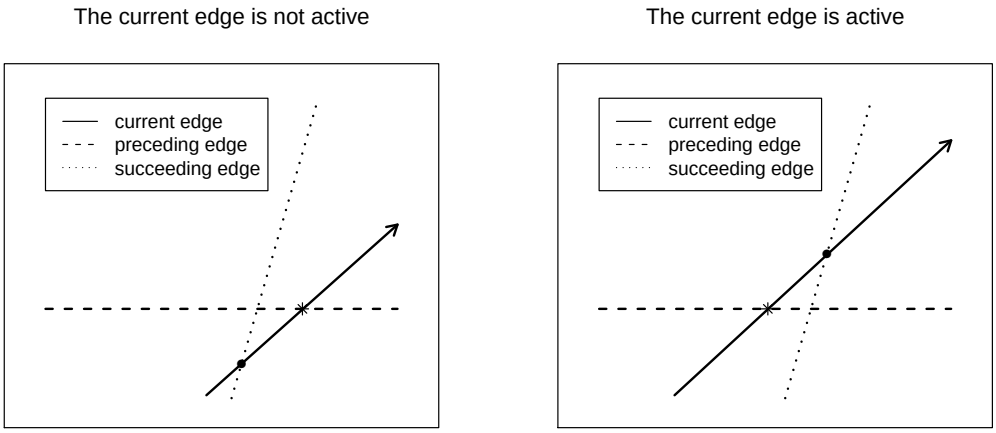

Figure 6.2: *The asterisk point is vprec and the dot point is vnext. In the left panel, the direction from asterisk to dot is opposite to the direction of the current edge, thus the current edge is not active and should be eliminated. While in the right panel, the direction from asterisk to dot is consistent to the direction of the current edge, the current edge is active.*

if the current edge was validated—or moves back to the preceding edge. The edges, succeeding or preceding, are always picked in the circular order from the yet non-eliminated edges. If the algorithm moves backward, then the current edge is eliminated and the pointers to the next and preceding edges are updated, to bypass the invalidated edge in the circular list. If the eliminated edge happens to be marked as terminal, then the pointer `etrm` is updated as well: the terminal status is passed to its preceding edge—unless the toggle `trms` is turned on; in such a case it is passed to the succeeding edge.

The algorithm terminates after accomplishing the full round and validates the initial/final edge. Reaching the initial edge means that `ecur` become to `etrm`, and `invl` is turned off—the edge mush be reached by the *forward*, not *backward* motion. The toggle `trms` is then turned on, and if the edge is successfully validated, the algorithm terminates and outputs the collection of the incoming vertices of the active

edges, in their circular order. The last step to obtain the envelope is to connect all the vertices in their circular order.

The algorithm may terminate also declaring the intersection of the halfspaces empty. This happens either when the angle of two neighboring vertices is found not less than π in the counter-clockwise direction, or when the number of active edges becomes smaller than three. The former situation may happen when the directions we selected are contained in a halfspace with boundary line passing through the origin. the latter may happen when p is two large, for instance p is greater than the halfspace depth of the Tukey median, see Theorem 4.2.1.

To sum it up, we will restate the entire algorithm. The notations we are going to use are defined as following:

`e` - the edge;

`v` - the vertex;

`ecur` - the current edge;

`eprec` - the preceding edge;

`enext` - the succeeding edge;

`etrm` - the initial/final edge;

`a` - the angle, in $(0, 2\pi]$, between `ecur` and `enext`;

`vprec` - the outgoing vertex, the intersection of its incoming edge (`ecur`) and its outgoing edge (`eprec`);

`vnext` - the incoming vertex, the intersection of its incoming edge (`enext`) and its outgoing edge (`ecur`);

`nact` - the number of active edges;

`invl` - the toggle to determine the direction, backward or forward, of the algorithm motion;

trms - the toggle to determine whether to stop the algorithm or not.

The input is a collection of circularly ordered N initial edges and the output of the algorithm is a collection of circularly ordered vertices; where the intersection envelope can be constructed. After the initialization, the algorithm repeats the validation procedure to eliminate non-active edges until all remaining edges are active or the intersection of the halfspaces is empty. Each vertex is linked to its incoming edge. For example vprec is linked to ecur and vnext is linked to enext. We use invl$= 1$ to represent invl is turned on, moving backward and invl$= 0$ to represent invl is turned off, moving forward. We use trms$= 1$ to represent trms is turned on, stopping the algorithm and use trms$= 0$ to represent trms is turned off, not stopping the algorithm.

Step 1: Initialization;

- Set up nact $= N$, the total number of input edges. Turn on the toggles invl by setting them as one. Turn off trms by setting it as zero. Pick up any edge as the current edge, ecur, and set its preceding edge as the preceding edge, eprec, its succeeding edge as the succeeding edge, enext. Note this is done in counter-clockwise order. Set up etrm as the current edge.

- Compute the vertex vprec as the intersection of eprec and ecur and the vertex enext as the intersection of ecur and enext. Also compute the angle a between ecur and enext.

Step 2: Validation;

- If the direction from the vertex vprec to the vertex vnext is consistent with

the direction of the current edge, and if the toggle `invl` is turned off and `ecur` is identical to `etrm` then turn on `trms` and go to **Step 4** otherwise turn off the toggle `invl`.

- If either the direction from the vertex `vprec` to the vertex `vnext` is not consistent with the direction of the current edge or `vprec` is identical to `vnext`, then turn on the toggle `invl`.

Step 3: Elimination;

- If the toggle `invl` is turned off, move forward one edge by setting the succeeding edge as `ecur`, the current edge as `eprec`, and the edge next to the succeeding edge in counter-clockwise order as `enext`. Set `vprec` as `vnext` and compute `vnext` as the intersection of `ecur` and `enext`. Compute the angle `a` between `ecur` and `enext`.

- If the toggle `invl` is turned on, set `etrm` as `eprec` if `etrm` is identical to `ecur`, eliminate the current edge and set up $\texttt{nact} = \texttt{nact} - 1$. Move backward one edge by setting the preceding edge as `ecur`, the edge next to the preceding edge in clockwise order as `eprec`, and the current edge as `enext`. If `etrm` is identical to `ecur`, set `vprec` as the intersection of `eprec` and `ecur`, otherwise set `vprec` as the vertex linked to `ecur`. Set `enext` as the intersection of `ecur` and `enext`. Compute the angle `a` between `ecur` and `enext`.

- If `nact` is less than 3 or `a` is no less than π, turn on `trms` by setting it as one. Stop the algorithm and return warning "The envelope is empty". Otherwise carry on and go to **Step 2** repeating the validation procedure.

Step 4: Construction;

- This last step will construct the envelope by connecting all vertices linked to active edges.

The crucial part of the algorithm is to eliminate non-active edges according to the results from the validation procedure. For this reason we call it the elimination algorithm of envelopes of directional quantile lines. All the edges can be validated by one pass although some edges may be checked more than once. For circularly ordered N initial edges, the time and storage complexity of the just described algorithm is $O(N)$. In Figure 6.3, we illustrate the envelope of $.10$-th directional quantile lines using Nepali children height and weight data. All the lines are 100 pre-selected directional quantile lines. The grey lines are non-active lines and the dark lines are active lines validated by the elimination algorithm. The inside envelope is the envelope of $.10$-th directional quantile lines.

6.3 Halfspace depth contours via directional quantiles

Theorem 4.2.1 says that the p-th directional quantile envelope is equivalent to the upper level set of depth: $D(p) = \{x : \ d(x) \geq p\}$. The elimination algorithm described in the last section is the core of the fast algorithm for approximate computation of halfspace depth contours via directional quantiles. The simplest strategy is to take directions spaced uniformly. That is, we add a pre-step before our core elimination algorithm:

Step 0: Generation;

- Generate d uniformly spaced directions in the interval $(-\pi, \pi)$. Compute the directional quantile lines corresponding to the d selected directions.

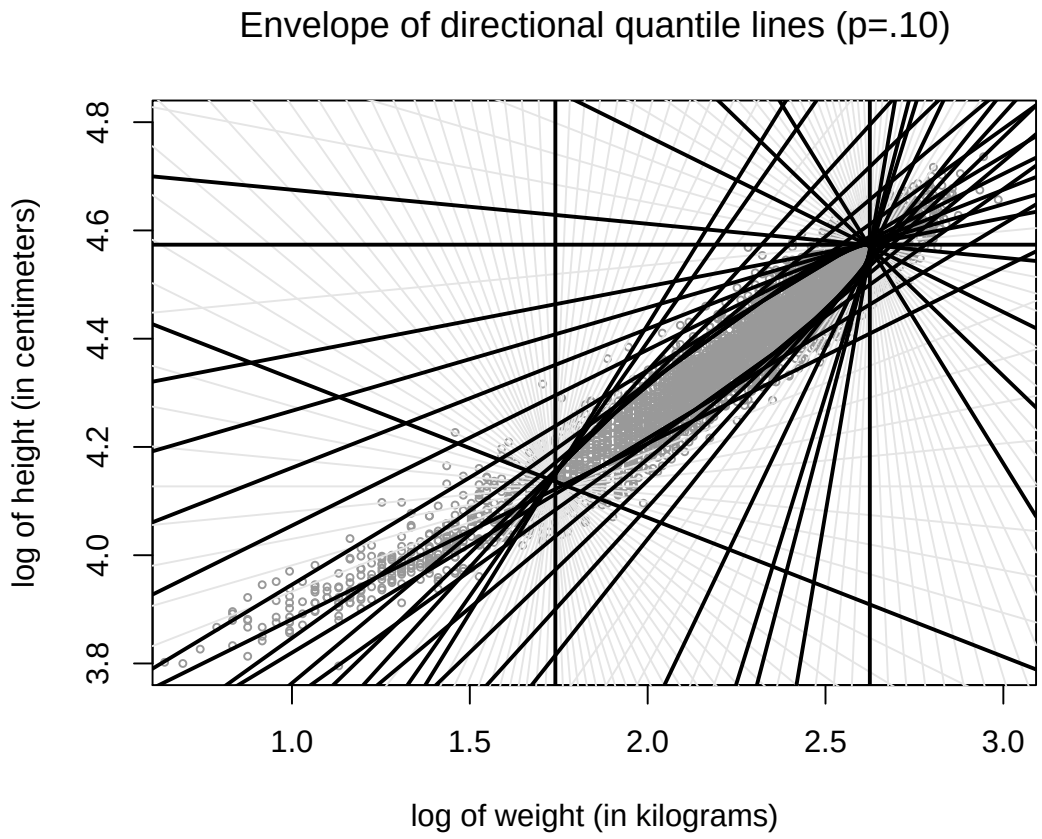

Figure 6.3: *All the lines are* 100 *pre-selected directional quantile lines. The grey lines are non-active lines and the dark lines are active lines validated by the elimination algorithm. The inside envelope is the envelope of* $.10$*-th directional quantile lines.*

In Figure 6.3, we used $d = 100$ uniformly spaced directions. The preselected number d determines how many lines we have before we start the elimination. It affects the accuracy of the approximation. In Figure 6.4 we illustrate two approximations of $.10$-th envelope with 50 uniformly spaced directions, in the left panel, and 500 uniformly spaced directions, in the right panel. The right one is smoother than the left one, however the difference is fairly minor and the running time is still reasonable even for $d = 1000$. In practice, the selection of the number of directions really depends on the purpose. If it is just for illustration, our experience is that 5% of the sample size will be enough.

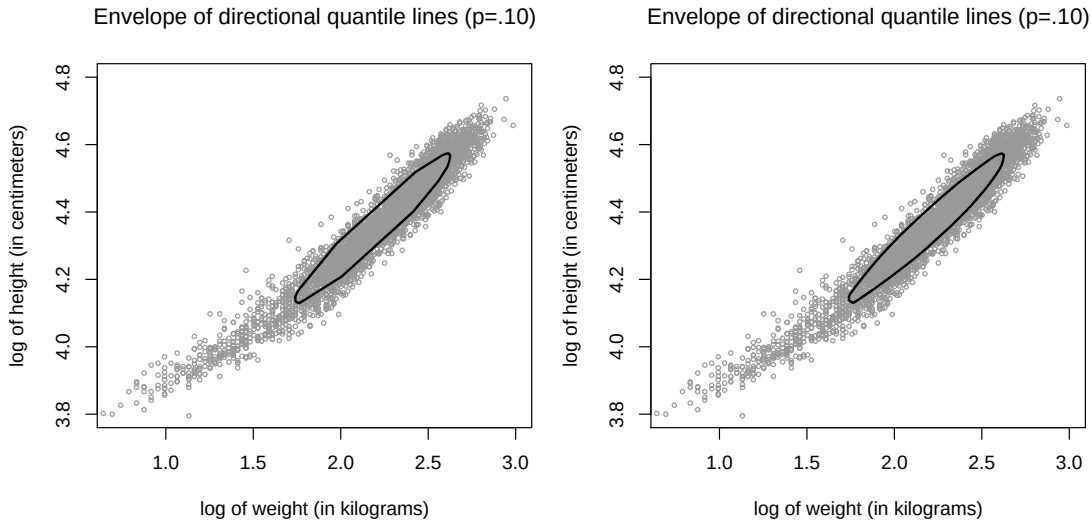

Figure 6.4: *In the left panel there are* 50 *uniformly spaced directions, while in the right penal there are* 500 *uniformly spaced directions. The right one is smoother than the left one, however the difference is fairly minor.*

Another way to improve the accuracy of the approximation is to select directions according to principal component analysis. The strategy is to select more around the directions of the first principal component. This can be also observed in Figure 6.3, where we have more active lines around the two corners. However, we prefer the method of uniformly spaced directions rather than any other fancy methods. The reason is we can always reach the same accuracy by increasing the number of directions. Given the speed of the elimination algorithm, there is really no need to find any alternative method for selecting directions.

What we provide here is an approximate algorithm to compute halfspace depth contours via directional quantiles. Some theoretical parts, like the relationship between accuracy and the number of directions, the relationship between accuracy and how to optimally select directions, the implementation in exact arithmetics, the existing (albeit alow) exact approaches, etc., are beyond the scope of this thesis.

Chapter 7

Halfspace depth and directional regression quantiles

7.1 Halfspace depth contours

The halfspace depth of a point $x \in \mathbb{R}^d$ with respect to a probability distribution P is $d(x, P) = \min_{\|u\|\neq 0} P\{y : \ u^T y \geq u^T x, \ y \in \mathbb{R}^d\}$. Its empirical version is obtained by substituting P_n for P. That is, $d(x, P_n) = \inf_{\|u\|\neq 0} \#\{i : \ u^T y_i \geq u^T x, \ y_i \in \mathbb{R}^d\}/n$, where $\#\{\cdot\}$ denotes the number of data points in $\{\cdot\}$. The p-th depth contour D_p is the collection of $x \in \mathbb{R}^d$ such that $d(x) \geq p$. The depth, $d(x)$, of x is affine equivariant and D_p is convex and nested, as described in Section 2.5.

Theorem 4.2.1 says that the p-th directional quantile envelope is equivalent to the p-th depth contour. Theorem 3.3.4 implies that the closed form of p-th depth contour of an elliptically symmetric distribution P of a random variable X is

$$\{x : \ (x-\mu)^T \Sigma^{-1} (x-\mu) \leq c_p^2\},$$

where c_p is a constant such that $\mathbb{P}[s^T \Sigma^{-1/2} (X-\mu) \leq c_p] = p$ with $s \in \mathbb{S}^{d-1}$. Therefore, any tangent line of the p-th depth contour D_p of an elliptically symmetric distribution P splits the space into two halfspaces with masses p and $1-p$, respectively. Theorem 3.3.4 is a special case of Corollary 4.3.1, where the assumption is that the contour is smooth. Theorem 4.3.1 gives more general conditions: the tangent is unique and the

maximum mass on a hyperplane is zero, i.e. $\Delta(P)$=0. However, the interpretation of depth contours in empirical distribution case has not been studied in the literature yet. Also the conditions of Theorem 4.3.1 are difficult to verify in practice. Therefore, in this section, we will continue to study the problem of interpreting depth contours.

7.1.1 Discrete data points

We will start our study from empirical distributions, the case with discrete data points. Note that any depth contour for discrete data points is a convex polygon and on its vertex (extremal point) there are more than one support halfspace; see Theorem 4.3.1. We will first study the structure of depth contours in two dimensional space. Suppose we have n data points; the p-th depth contour D_p is the intersection of all the halfspaces containing at least $n - \lceil np \rceil + 1$ data points, which can be formalized as the following proposition (He and Wang, 1997; Rousseeuw and Ruts, 1999).

Proposition 7.1.1 *In discrete data points case, for any $0 < p < 1$, we have*

$$D_p = \bigcap_{H \in \mathcal{H}(m)} H, \quad m = n - \lceil np \rceil + 1,$$

where $\mathcal{H}(m)$ is the collection of all closed halfspaces containing at least m data points and $\lceil x \rceil$ denotes the smallest integer which is greater than x.

Obviously, if p is greater than the depth of the Tukey median, then D_p and thus $\bigcap_{H \in \mathcal{H}(m)} H$ is empty. As a consequence of Proposition 7.1.1, if p_1 and p_2 satisfy $\lceil np_1 \rceil = \lceil np_2 \rceil$, then $D_{p_1} = D_{p_2}$. If there are no three points lying on a line, then the p-th depth contour D_p is the intersection of all the halfspaces containing exact $n - \lceil np \rceil + 1$ data points. The boundary of the p-th depth contour ∂D_p is the union of all the intersections of D_p and the boundary of those halfspaces containing at least $n - \lceil np \rceil + 1$ data points.

Proposition 7.1.2 *The boundary of the p-th depth contour D_p, $0 < p < 1$, can be written as*

$$\partial D_p = \bigcup_{H \in \mathcal{H}(m)} \left(D_p \bigcap \partial H \right), \quad m = n - \lceil np \rceil + 1,$$

where $\mathcal{H}(m)$ is defined as in Proposition 7.1.1.

Proof. Suppose $x \in \partial D_p$, then there must be a halfspace $H \in \mathcal{H}(n - \lceil np \rceil + 1)$ such that $x \in \partial H$ from Proposition 7.1.1. Otherwise, x is an inner point of any halfspace in $\mathcal{H}(n - \lceil np \rceil + 1)$ and thus an inner point of D_p. We have $x \in D_p \bigcap \partial H$, that is, $D_p \subset \bigcup_{H \in \mathcal{H}(m)} (D_p \bigcap \partial H)$ with $m = n - \lceil np \rceil + 1$.

On the other hand, if $x \in D_p \bigcap \partial H$ for some $H \in \mathcal{H}(n - \lceil np \rceil + 1)$, we have $x \in D_p$ and $x \in \partial H$. If x is an inner point of D_p, then x must be an inner point of H, which is contradictory to $x \in \partial H$. $x \in \partial D_p$ and then $\bigcup_{H \in \mathcal{H}(m)} (D_p \bigcap \partial H) \subset D_p$ with $m = n - \lceil np \rceil + 1$. □

If there do not exist three or more points lying on a line, the halfspaces in Proposition 7.1.2 can be replaced be those containing exact $n - \lceil np \rceil + 1$ data points. That is, $\partial D_p = \bigcup_{H \in \mathcal{G}(m)} (D_p \bigcap \partial H), \quad m = n - \lceil np \rceil + 1$, where $\mathcal{G}(m)$ is the collection of all closed halfspaces containing m data points.

Proposition 7.1.3 *Suppose that $0 < p_1$ and $0 < p_2$ are no greater than the depth of the Tukey median and satisfy $\lceil np_2 \rceil - \lceil np_1 \rceil = 1$. Then we have*

$$d(x) = \lceil np_1 \rceil / n, \quad \textit{given} \quad x \in D_{p_1} / D_{p_2}$$

and $\partial D_{p_1} \subset D_{p_1} / D_{p_2}$, if there are no three points lying on a line.

Proof. Suppose that $x \in D_{p_1} / D_{p_2}$, then $p_1 \leq d(x) < p_2$ and $np_1 \leq nd(x) < np_2$. Since $nd(x)$ is an integer and $\lceil np_2 \rceil - \lceil np_1 \rceil = 1$, the only possible value for $nd(x)$ is $\lceil np_1 \rceil$ considering that $\lceil np_1 \rceil \leq nd(x) < \lceil np_2 \rceil$. Therefore, we have $d(x) = \lceil np_1 \rceil / n$.

To prove that $\partial D_{p_1} \subset D_{p_1}/D_{p_2}$, it is sufficient to prove that $\partial D_{p_1} \bigcap D_{p_2} = \emptyset$. Suppose that $\partial D_{p_1} \bigcap D_{p_2} \neq \emptyset$ and let $x \in \partial D_{p_1} \bigcap D_{p_2}$. By Proposition 7.1.1, we have x belongs to any halfspaces containing exact $n - \lceil np_1 \rceil + 1$ or $n - \lceil np_2 \rceil + 1$ data points. As $x \in \partial D_{p_1}$, there exists a halfspace H_1 containing exact $n - \lceil np_1 \rceil + 1$ data points such that $x \in \partial H_1$. Since $\lceil np_2 \rceil - \lceil np_1 \rceil = 1$, we can find a halfspace H_2 containing exact $n - \lceil np_2 \rceil + 1$ data points such that $H_1 \subset H_2$. Obviously $x \notin H_2$ because of $x \in \partial H_1$, which is contradictory to that x belongs to any halfspaces containing exact $n - \lceil np_2 \rceil + 1$ data points. Hence $\partial D_{p_1} \bigcap D_{p_2} = \emptyset$ and $\partial D_{p_1} \subset D_{p_1}/D_{p_2}$. Consequently, if $x \in \partial D_{p_1}$ then $d(x) = \lceil np_1 \rceil / n$. □

After studying the structure of depth contours and their boundary, we are now ready to explore the interpretation of depth contours of discrete data points.

Theorem 7.1.1 *Suppose that we have n data points in $\mathbb{R}$. For $0 < p < 1$, any boundary line of the p-th depth contour D_p splits the whole space $\mathbb{R}$ into two halfspaces such that in the closed halfspace containing D_p there are at least $n - \lceil np \rceil + 1$ data points and in its complement halfspace there are at most $\lceil np \rceil - 1$ data points. Particularly, if there do not exist three or more points lying on a line, the two closed halfspaces have exact $n - \lceil np \rceil + 1$ and $\lceil np \rceil - 1$ data points, respectively.*

Proof. Suppose l is a boundary line of the p-th depth contour D_p and H is the closed halfspace containing D_p with $\partial H = l$. The intersection of D_P and ∂H is a segment of l. Let x_1 be an inner point of this segment, we can find H_1 containing at least $n - \lceil np \rceil + 1$ data points such that $x_1 \in \partial H_1$ according to Proposition 7.1.2. As $D_p \in H_1$ and x_1 is not an end point of the boundary, we have $H_1 = H$. Therefore H contains at least $n - \lceil np \rceil + 1$ data points; and its complement halfspace H^c contains at most $\lceil np \rceil - 1$ data points. Particularly, if there do not exist three or more points lying on a line, halfspace H_1 and hence H will contain exact $n - \lceil np \rceil + 1$ data points.

□

In the case where there are three points lying on a line, we are still able to find how exactly the boundary line of depth contour splits the space, even in dimensions greater than two. This is guaranteed by the result of Struyf and Rousseeuw (1999): halfspace depth characterizes the empirical distribution. Similarly, if an atomic probability distribution has finite support, we will be able to retrieve all the probabilistic information from depth contour by the result of Koshevoy (2002) and Hassairi and Regaieg (2007) that halfspace depth characterizes an atomic probability distribution with finite support.

7.1.2 Absolutely continuous distributions

Discrete data points are always sampled from a population distribution. It is important to study the interpretation of depth contours of population distributions. It may be difficult to look at general population distributions. We will focus our attention on absolutely continuous population distributions, by which we will mean population distribution absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. For example, an elliptically symmetric distribution with nonsingular Σ is absolutely continuous and so does a normal distribution with nonsingular Σ. Similar to the discrete data points case, the p-th depth contour D_p.

Proposition 7.1.4 *For any $0 < p < 1$, suppose that D_p is the p-th depth contour of a probability distribution P in $\mathbb{R}^d$, we have*

$$D_p = \bigcap_{H \in \mathcal{H}(1-p)} H,$$

where $\mathcal{H}(1-p)$ is the collection of all closed halfspaces with probability measure more than $1-p$. If P is absolutely continuous with positive density almost everywhere, then

$$D_p = \bigcap_{H \in \mathcal{G}(1-p)} H,$$

where $\mathcal{G}(1-p)$ is the collection of all closed halfspaces with probability measure $1-p$.

Proof. The first part of the proposition is a direct consequence of the definition of halfspace depth and depth contours. That is, $d(x) = \min_{\|u\|\neq 0} P\{y:\ u^T y \geq u^T x,\}$ and $D_p = \{x:\ d(x) \geq p\}$; see Rousseeuw and Ruts (1999). In the second part, if P has positive density almost everywhere, then in any direction $\|u\| \in \mathbb{S}^{d-1}$, we have a unique closed halfspace $H = \{x : u^T x \geq c\}$ such that $P\{H\} = 1 - p$. Suppose that $x_0 \in D_p$; then $d(x_0) \geq p$ and for any direction u, we have $P\{x : u^T x \geq u^T x_0\} \geq p$. Apparently $\{x : u^T x \geq u^T x_0\} \subset \{x : u^T x \geq c\}$, which means $x_0 \in \{x : u^T x \geq c\}$. That is, $D_p \subset \bigcap_{H\in\mathcal{G}(1-p)} H$. If $x_0 \in \bigcap_{H\in\mathcal{G}(1-p)} H$, for any direction u and c satisfying $P\{x : u^T x \leq c\} = 1 - p$, then $x_0 \in \{x : u^T x \leq c\}$. Hence $u^T x_0 \leq c$ and $P\{x : u^T x \leq u^T x_0\} \leq P\{x : u^T x \leq c\} = 1 - p$. For any direction u, $P\{x : u^T x \geq u^T x_0\} \geq p$. Therefore $d(x_0) \geq p$ and $x_0 \in D_p$. □

In Proposition 7.1.4, the condition of absolute continuity with positive density almost everywhere can be relaxed. For example, it can be replaced by the condition that the probability measure of any hyperplane is zero and any nonzero Lebesgue measure set has nonzero probability. As a result, for any closed halfspace H, $P\{\partial H\} = 0$; and for any two closed halfspaces H_1 and H_2 with parallel boundary and $H_1 \subset H_2$, $P\{H_2/H_1\} \neq 0$. The latter requirement is sharp; for instance, let P be a probability distribution of $0.5U[-2,-1] + 0.5U[1,2]$, where $U[a,b]$ denotes the continuous uniform distribution in the interval $[a,b]$. We have $D_{0.5} = [-1,1]$ while the intersection of all halflines with probability 0.5 is empty. We can avoid this situation by replacing $\mathcal{G}(1-p)$ by $\tilde{\mathcal{G}}(1-p)$, where $\tilde{\mathcal{G}}(1-p) = \big\{H_u = \{x : u^T x \geq c_u\} : c_u = \sup\{c : P\{x : u^T x \geq c\} = 1 - p\}\big\}$.

Proposition 7.1.5 *For any $0 < p < 1$, suppose that D_p is the p-th depth contour of a probability distribution P in $\mathbb{R}^d$. If P is absolutely continuous and Lebesgue measure is*

absolutely continuous with respect to P, then

$$\partial D_p = \bigcup_{H \in \mathcal{G}(1-p)} \left(D_p \bigcap \partial H \right),$$

where $\mathcal{G}(1-p)$ is defined as in Proposition 7.1.4.

Proof. Suppose $x_0 \in \partial D_p$, then $x_0 \in D_p$ and there exists a halfspace $H \in \mathcal{G}(1-p)$ such that $x_0 \in \partial H$. Otherwise if x_0 is an inner point of all closed halfspaces in $\mathcal{G}(1-p)$, then x_0 will be an inner point of D_p according to Proposition 7.1.4. Therefore, $x_0 \in D_p \bigcap \partial H$, i.e. $\partial D_p \subset \bigcup_{H \in \mathcal{G}(1-p)} (D_p \bigcap \partial H)$. If there exists a halfspace $H \in \mathcal{G}(1-p)$ such that $x_0 \in D_p \bigcap \partial H$, then $x_0 \in \partial D_p$. Otherwise, if $x_0 \in D_p^\circ$, the inner of D_p, we can find $x_0 - \lambda u \in D_p$ with λ is small enough where u is the normal direction of H. Since $x_0 \in \partial H$ we have $x_0 - \lambda u \notin H$ and hence $x_0 - \lambda u \notin D_p$. Therefore $\partial D_p \supset \bigcup_{H \in \mathcal{G}(1-p)} (D_p \bigcap \partial H)$. □

Even when P does not have positive density almost everywhere, the proposition still holds true, we replace $\mathcal{G}(1-p)$ by $\tilde{\mathcal{G}}(1-p)$.

Proposition 7.1.6 *Suppose that P is an absolutely continuous distribution in $\mathbb{R}^d$. Then for any $0 < p < 1$,*

$$d(x_0) = p \quad \textit{if} \quad x_0 \in \partial D_p, \quad \textit{and} \quad \partial D_p \subset D_p / D_p^-,$$

where $D_p^- = \{x : d(x) > p\}$. If P has positive density almost everywhere, then $D_p^- = D_p^\circ$ and $\partial D_p = D_p / D_p^\circ$, where D_p° is the inner of D_p.

Proof. We first prove that if $x_0 \in \partial D_p$, then $d(x_0) = p$. According to the definition of depth contour, $x_0 \in \partial D_p$ means $d(x_0) \geq p$. If $d(x_0) = p_0 \neq p$, then $x_0 \in D_{p_0} = \bigcup_{H \in \mathcal{H}(1-p_0)} H$ according to Proposition 7.1.4. As P is absolutely continuous, x_0 is inner point of any halfspace in $\mathcal{H}(1-p)$ and thus an inner point of

D_p. Therefore $d(x_0) = p$. As a consequence, $\partial D_p \subset D_p/D_p^-$ and $D_p^- \subset D_p^\circ$. If P has positive density almost everywhere, and x_0 is an inner point of D_p°/D_p^-, then $d(x_0) = p$. We can find a halfspace $H = \{x : u^T x \geq u^T x_0\}$ with probability $1 - p$ such that $x_0 \in H$. Since x_0 is an inner point of D_p°/D_p^-, we can find a data point $x_0 - \lambda u \in D_p^\circ/D_p^-$ with $\lambda > 0$ small enough such that $d(x_0 - \lambda u) = p$. However, $P\{x : u^T x \geq u^T(x_0 - \lambda u)\} > P\{x : u^T x \geq u^T x_0\} = 1 - p$, so $d(x_0 - \lambda u) < p$. That is a contradiction. Therefore $D_p^\circ/D_p^- = \emptyset$ and $\partial D_p = D_p/D_p^\circ$. □

Theorem 7.1.2 *Suppose that P is an absolutely continuous probability measure with positive density almost everywhere in $\mathbb{R}^d$. For any $0 < p < 1$, any tangent hyperplane of the p-th depth contour D_p divides $\mathbb{R}^d$ into two halfspaces. If the tangent hyperplane is unique, then the halfspace completely containing D_p has probability measure $1 - p$ and the other halfspace has p.*

Proof. Let l be a tangent hyperplane of D_p; l splits $\mathbb{R}^d$ into two halfspaces. Denote the halfspace completely containing D_p by H. To avoid any confusion, we assume H is closed. l passes through at least one point of D_p on ∂D_p, denoting it by x_0. According to Proposition 7.1.5, we know $\partial D_p = \bigcup_{H \in \mathcal{G}(1-p)} (D_p \bigcap \partial H)$. $x_0 \in \partial D_p$ means there exists a halfspace $H_1 \in \mathcal{G}(1 - p)$ such that $P \in D_p \bigcap \partial H_1$ and then $x_0 \in \partial H_1$. Proposition 7.1.4 says $D_p = \bigcap_{H \in \mathcal{G}(1-p)} H$. Since $H_1 \in \mathcal{G}(1 - p)$, $D_p \subset H_1$. Therefore, we have $D_p \subset H_1$ and $x_0 \in \partial H_1$. As a result, ∂H_1 is a tangent hyperplane of D_p passing through x_0. From the uniqueness of the tangent hyperplane, we know $l = \partial H_1$ and then $H = H_1$ considering both halfspaces contain D_p. Therefore $P\{H\} = P\{H_1\} = 1 - p$. As P is absolutely continuous, we obtain $P\{H^c\} = p$. □

Even when P does not have positive density almost everywhere, Theorem 7.1.2 still holds true. The proof is similar while we need to replace $\mathcal{G}(1 - p)$ by $\tilde{\mathcal{G}}(1 - p)$. Let

$H_{y,u}$ stand for the closed halfspace that passes through y with direction u pointing inside, i.e. $H_{y,u} = \{x : u^T(y - x) \geq 0\}$. Denote the Tukey median by μ_T, then $\mu_T = \arg\max_y \inf_{\|u\|=1} P(H_{y,u})$. If P is absolutely continuous, then for any $p \in (0, d(\mu_T)]$, there exists at least one point $x \in \mathbb{R}^d$ such that $d(x) = p$; see Massé (2004). The following proposition gives a condition when the Tukey median is unique; similar condition can be found in Massé and Theodorescu (1994) and Mizera and Volauf (2001).

Proposition 7.1.7 *Suppose $P \in \mathbb{R}^d$ is an absolutely continuous probability measure with positive density almost everywhere, then its Tukey median μ_T is unique.*

Proof. Suppose u_T is the direction maximizing $\inf_{\|u\|=1} P(H_{\mu_T,u})$. If there exists another point $\tilde{\mu}_T$ having the maximum halfspace depth, then we will have three situations: $\tilde{\mu}_T \in H^\circ_{\mu_T,u_T}$, $\tilde{\mu}_T \in \partial H_{\mu_T,u_T}$, and $\tilde{\mu}_T \in H^\circ_{\mu_T,-u_T}$, where H° stands for the interior part of H.

If $\tilde{\mu}_T \in H^\circ_{\mu_T,u_T}$, we have the Lebesgue measure of $H_{\mu_T,u_T}/H_{\tilde{\mu}_T,u_T}$ is greater than zero. As P has positive density almost everywhere, we have $P\left(H_{\mu_T,u_T}/H_{\tilde{\mu}_T,u_T}\right) \neq 0$. Thus, $P(H_{\mu_T,u_T}) > P(H_{\tilde{\mu}_T,u_T})$, which means $D(\mu) > D(\tilde{\mu})$. This is a contradiction.

If $\tilde{\mu}_T \in \partial H_{\mu_T,u_T}$, then any point $\mu(\lambda) = \tilde{\mu}_T - \lambda(\tilde{\mu}_T - \mu_T)$ with $\lambda \in [0, 1]$ is a Tukey median according to the convexity of depth contours. u_T is a direction on which $\mu(\lambda)$ obtains its maximum depth. The sequence $\mu(0.5) - \delta u_T$ tends to $\mu(0.5)$ as δ tends to zero. Denote u_T^δ as a direction maximizing the depth of $\mu(0.5) - \delta u_T$ such that u_T^δ tends to u_T as δ tends to zero. As a result we either obtain $u_T^\delta = u_T$ or $\mu(0.5) \in H_{\mu(0.5)-\delta u_T, u_T^\delta}$ when δ is small enough. Any of them will derive a contradiction.

In the last case, if $\tilde{\mu}_T \in H^\circ_{\mu_T,-u_T}$, then there does not exist a suitable direction to maximize its halfspace depth of $\mu(0.5)$. It is a contradiction again. □

Absolutely continuity by itself does not guarantee the uniqueness of the Tukey median. For instance, the probability distribution corresponding to $0.5U[-2,-1] + 0.5U[1,2]$ is absolutely continuous, where $U[a,b]$ denotes the continuous uniform distribution in the interval $[a,b]$. However, any point in the interval $[-1,1]$ is a Tukey median. Any elliptically symmetric distribution with nonsingular Σ has unique Tukey median; the Tukey median is its center μ and the depth is 0.5. More generally, any angularly symmetric distribution has its center as the unique Tukey median and the depth is 0.5; see Rousseeuw and Ruts (1999).

7.1.3 Continuity of halfspace depth

Continuity of halfspace depth has been studied by many researchers under various conditions. For example, Massé (2002) pointed out that $d(x)$ was continuous if $\Delta(P) = 0$. Massé and Theodorescu (1994) proved that if P was absolutely continuous with connected support, then D_p was the closure of D_P^- and the Tukey median is unique. Mizera and Volauf (2001) verified that these were true if P had contiguous support; see Definition 3.1.3. Rousseeuw and Ruts (1999) also studied the continuity of halfspace depth and proved the Ray Basis Theorem and Inverse Ray Basis Theorem on Tukey median. For $0 < p$, the following conditions are the most popular ones.

C1: $\Delta(P) = 0$. That is, for any $c \in \mathbb{R}$ and $u \in \mathbf{S}^{d-1}$, $P\{x : u^T x = c\} = 0$.
C2: $D_p = \overline{D_p^-}$, where $D_p^- = \{x : \ d(x) > p\}$ and $\overline{D_p^-}$ is the closure of D_p^-.
C3: The Tukey median is a singleton, i.e. unique.
C4: D_p is smooth. That is, for any $x \in \partial D_p$, its tangent halfspace is unique.

C1 is a basic condition; many continuous results are derived under it. For example, $d(x)$ was continuous if **C1** is satisfied (Massé, 2002). **C2** and **C3** are satisfied if **C1** is

satisfied and P is contiguous (Mizera and Volauf, 2001). **C3** is an important property, which has been derived by Massé and Theodorescu (1994), Mizera and Volauf (2001), and Proposition 7.1.7. The three situations are included in the following proposition, which gives a sufficient condition of **C3** that is the most general condition in the literature. Given a convex set $A \in \mathbb{R}^d$, its ε-neighborhood is defined as $A_\varepsilon = \{x : \inf_{y \in A} d(x, y) < \varepsilon\}$, where $d(\cdot, \cdot)$ is the usual Euclidean distance.

Proposition 7.1.8 *Suppose P satisfies condition* **C1**. *If for any $a < b$ and $\varepsilon > 0$, $A = \{x : a < u^T x < b\} \cap \{x : \ d(x) = d(\mu_T)\}_\varepsilon \neq \emptyset$ where μ_T has maximum depth, we have $P(A) > 0$, then condition* **C3** *are met.*

The proof is similar to that of Proposition 7.1.7 and will not provided. The depth $d(x)$ is non-increasing along a ray starting from μ_T as D_p is nested and continuous if **C1** is satisfied. The following proposition gives a condition such that $d(x)$ is strictly decreasing.

Theorem 7.1.3 *Suppose that P satisfies conditions* **C1**, **C2** *and* **C3**. *For any $u \in \mathbb{S}^{d-1}$ and $a \geq 0$, the map*

$$a \longmapsto d(au + \mu_T)$$

is continuous and non-increasing. Particularly, if $a \in [0, a^*]$ then the map is strictly decreasing, where $a^* = \inf_{a \geq 0}\{a : d(au + \mu_T) = 0\}$.

Proof. The continuity is a direct consequence of Massé (2002). Let $a_1 > a_2$, we need to prove that $d(a_1 u + \mu_T) \leq d(a_2 u + \mu_T)$. Suppose that v is a minimal direction of $a_2 u + \mu_T$, that is $P\left(H_{a_2 u + \mu_T, v}\right) = d(a_2 u + \mu_T)$. We would have the following three situations.

(i) $u^T v > 0$: If $u^T v > 0$, then $a_1 u + \mu_T \in H_{a_2 u + \mu_T, v}$ and $H_{a_1 u + \mu_T, v} \subseteq H_{a_2 u + \mu_T, v}$. As a sequence, $d(a_2 u + \mu) = P\left(H_{a_2 u + \mu_T, v}\right) \geq P\left(H_{a_1 u + \mu_T, v}\right) = d(a_1 u + \mu)$.

(ii) $u^T v = 0$: If $u^T v = 0$, then $\mu_T \in H_{a_2 u+\mu_T, v}$ and $d(\mu_T) \leq P\left(H_{a_2 u+\mu_T, v}\right) = d(a_2 u + \mu_T)$. Since μ_T is the Tukey median, $d(\mu_T) \geq d(a_2 u + \mu_T)$ and $d(\mu_T) = d(a_2 u + \mu_T)$. As the Tukey median is unique, we have $a_2 u + \mu_T = \mu_T$ and thus $a_2 = 0$. Therefore, $d(a_2 u + \mu_T) = d\mu_T \geq d(a_1 u + \mu_T)$.

(iii) $u^T v < 0$: If $u^T v < 0$, then $\mu_T \in H_{a_2 u+\mu_T, v}$ and $d(a_2 u + \mu_T) = H_{a_2 u+\mu_T, v} \geq d(\mu_T)$. Therefore $a_2 u + \mu_T = \mu_T$ and $a_2 = 0$; and $d(a_1 u + \mu) \leq d(a_2 u + \mu)$.

Particularly, if $a_2 < a_1 < a^*$, where $a^* = \inf\{a \geq 0, d(au + \mu_T) = 0\}$, we claim that $d(a_2 u + \mu_T) > d(a_1 u + \mu_T)$. Otherwise suppose $d(a_2 u + \mu_T) = d(a_1 u + \mu_T)$. In the case $u^T v \leq 0$, we have derived that $a_2 = 0$, and then $a_1 > 0$. Notice that $d(a_1 u + \mu_T) = d(a_2 u + \mu_T) = d(\mu_T)$, we have $a_1 u + \mu_T$ is also the Tukey median, which is contradictory to the condition **C3**. On the other hand, if $u^T v > 0$, we will have $a_1 u + \mu_T \in H_{a_2 u+\mu_T, v}$, and $d(x) \leq p$ for any $x \in H_{a_2 u+\mu_T, v}$ where $p = d(a_2 u + \mu_T)$. As $d(x)$ is continuous under condition **C1**, we can find a sequence $x_n \to a_1 u + \mu_T$, such that $d(x_n) \to d(a_1 u + \mu_T)$. Considering that $d(a_1 u + \mu_T) < d(\mu_T)$ and P satisfies condition **C2**, we can make $d(x_n) > p$ for any n. If n is big enough, $\|x_n - (a_1 u + \mu_T)\| < \varepsilon$ where ε could be any small positive number, which implies that $x_n \in H_{a_2 u+\mu_T, v}$ as well. For these x_ns, $d(x_n) \leq p$, a contradictory to the fact $d(x_n) > p$ for any n. Therefore, $d(a_2 u + \mu_T) > d(a_1 u + \mu_T)$ if $a_2 < a_1 < a^*$. □

The condition of $a_2 < a_1 < a^*$ is necessary for the map $a \longmapsto d(au + \mu_T)$ to be strictly decreasing unless $a^* = \infty$. For instance, if $a^* < \infty$ and $a^* < a_2 < a_1$ we will have $d(a_2 u + \mu) = d(a_1 u + \mu) = 0$.

Theorem 7.1.4 *Suppose that P satisfies conditions* **C1**, **C2**, *and* **C3**. *Denote $D_p(u, c) = D_p \cap \{x : \ u^T x = c\}$. Then the sets $D_p(u, c)$ are closed and convex for any fixed u and c; and they are nested with respect to p. If P satisfies condition* **C4** *as well, $D_{p^*(u,c)}(u, c)$ is a singleton, where $p^*(u, c) = \sup_{p>0}\{p : \ D_p(u, c) \neq \emptyset\}$.*

Proof. The sets D_p are closed and convex; and they are nested with respect to p under the condition **C1**, **C2** and **C3**; see Rousseeuw and Ruts (1999). As a consequence, the system $D_p(u, c)$, formed by truncations of D_p by an one dimensional hyperplane $\{x : u^T x = c\}$, consists of closed convex sets which are nested as well. To prove that $D_{p^*(u,c)}(u, c)$ is a singleton, we need to prove that the hyperplane $\{x : u^T x = c\}$ is the support hyperplane of depth contour $D_{p^*(u,c)}$. Without loss of generality, we assume the Tukey median μ_T lies on the halfspace $\{x : u^T x \geq c\}$. We have that $\{x : u^T x = c\}$ is the support hyperplane, if for any point x_0 on the halfspace $\{x : u^T x < c\}$, its depth $d(x_0)$ is less than $p^*(u, c)$. Otherwise if $d(x_0) \geq p^*(u, c)$, then the segment line connecting x_0 and μ_T will intersect the hyperplane $\{x : u^T x = c\}$. Let us denote the intersection point by y_0. Notice that x_0 and y_0 can be written as $x_0 = a_1 u + \mu_T$ and $y_0 = a_2 u + \mu_T$, respectively, where $u = (x_0 - \mu_T)/\|x_0 - \mu_T\|$, $a_1 = \|x_0 - \mu_T\| > a_2 = \|y_0 - \mu_T\|$. According to Proposition 7.1.3, we have $d(y_0) > d(x_0) \geq p^*(u, c)$. However, the definition of $p^*(u, c)$ says that $d(y_0) \leq p^*(u, c)$, which is a contradiction. Hence for any point on the halfspace $\{x : u^T x < c\}$, its depth is less than $p^*(u, c)$. That is, $\{x : u^T x = c\}$ is the support hyperplane. From condition **C4**, we know that any support hyperplane is regular, which means that $\{x : u^T x = c\}$ is regular. Therefore, the intersection, $D_{p^*(u,c)}(u, c)$, of hyperplane $\{x : u^T x = c\}$ and depth contour $D_{p^*(u,c)}$ is a singleton. □

7.2 Directional regression quantiles

7.2.1 Linear programming for quantile regression

Definition 2.2.1 says: given a sample $(y_1, y_2, \cdots, y_n)$ of a response variable Y and the corresponding covariates $(x_1, x_2, \cdots, x_n)$ of explanatory variable X, where $X \in \mathbb{R}^d$, the p-th regression quantile (Koenker and Bassett, 1978) of Y conditioned by X is

obtained by solving

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \rho_p(y_i - x_i^T \beta),$$

where ρ_p is the "check" function $\rho_p(x) = x(p - I(x < 0))$. This quantile regression problem may be reformulated as a linear problem,

$$\begin{aligned} \text{minimize} \quad & p \sum_{i=1}^{n} e_i^+ + (1-p) \sum_{i=1}^{n} e_i^- \\ \text{subject to} \quad & x_i^T \beta^+ - x_i^T \beta^- + e_i^+ - e_i^- = y_i \\ & \beta^+, \beta^-, e_i^+, e_i^- \geq 0, \end{aligned} \tag{7.1}$$

where $\beta^+ = \beta \vee 0$, $\beta^- = (-\beta) \vee 0$, $e_i^+ = e_i \vee 0$, $e_i^- = (-e_i) \vee 0$, and $e_i = y_i - x_i \beta$ is the regression residual of the i-th observation. There are n equations and $2n + 2d$ unknowns in linear problem (7.1). If we call (7.1) a primal linear problem, then its dual linear problem is

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} b_i y_i \\ \text{subject to} \quad & p - 1 \leq b_i \leq p, \text{and} \sum_{i=1}^{n} x_i b_i = 0. \end{aligned} \tag{7.2}$$

A solution β and e_i is called feasible if it satisfies all the constrains in (7.1) for $i = 1, 2, \cdots, n$. Similarly, b_i, $i = 1, 2, \cdots, n$, is feasible if it satisfies the constrains in (7.2). The Weak Duality Theorem (Sposito,1975 and Nash and Sofer, 1996) says that if β, e_i, and b_i are feasible for (7.1) and (7.2), for $i = 1, 2, \cdots, n$, respectively, then $p \sum_{i=1}^{n} e_i^+ + (1-p) \sum_{i=1}^{n} e_i^- \geq \sum_{i=1}^{n} b_i y_i$. The Strong Duality Theorem says that if $p \sum_{i=1}^{n} e_i^+ + (1-p) \sum_{i=1}^{n} e_i^- = \sum_{i=1}^{n} b_i y_i$, then feasible solutions β, e_i, and b_i, $i = 1, 2, \cdots, n$, are optimal to (7.1) and (7.2), respectively. Sposito (1975) using the Complementary and Slackness Theorem derived a necessary and sufficient condition for feasible solutions β, e_i, and b_i, $i = 1, 2, \cdots, n$, to be optimal when $p = 0.5$. The following proposition gives a necessary and sufficient condition for feasible solutions to

be optimal for any $0 < p < 1$. The proof is similar to that in Sposito (1975) and will not be given here because of the space limitations.

Proposition 7.2.1 *Suppose that* (β, e_i) *and* b_i, $i = 1, 2, \cdots, n$, *are feasible solutions to* (7.1) *and* (7.2)*, respectively. Then they are optimal if and only if they satisfy the following two conditions:* $\sum_{i=1}^{n}(p - b_i)e_i^+ = 0$, *and* $\sum_{i=1}^{n}(1 - p + b_i)e_i^- = 0$.

Using Proposition 7.2.1, we can have the following proposition, which describes the relative position of observations to the p-th estimated quantile regression line.

Proposition 7.2.2 *Let* P *be the number of observations with positive residuals,* N *be the number of observations with negative residuals, and* Z *be the number of exactly fitted observations. If we include the intercept in the quantile regression, then*

$$P + N + Z = n$$

$$pP + (p - 1)N + \sum_{e_i^+ = e_i^- = 0} b_i = 0.$$

Proof. The first equation is obvious. We will just look at the second equation. From (7.1) and (7.2), we know $e_i^+ \geq 0$ and $p - b_i \geq 0$. The condition in Proposition 7.2.1 $\sum(p - b_i)e_i^+ = 0$ holds if and only if $(\tau - b_i)e_i^+ = 0$ for all i. If $e_i^+ > 0$ for some i, to make $(p - b_i)e_i^+ = 0$, we have $p - b_i = 0$, that is $p = b_i$. Similarly, if $e_i^- > 0$ for some i, we can obtain that $p - 1 = b_i$ because of the condition in Proposition 7.2.1, $\sum(1 - p + b_i)e_i^- = 0$. Actually, these can be also derived by directly applying the Complementary Slackness Theorem. Suppose that there are P observations with $e_i^+ > 0$, N observations with $e_i^- > 0$, and Z observations with $e_i^+ = e_i^- = 0$. In the quantile regression model, if we include the intercept, the constrain $\sum x_i b_i = 0$ in (7.2) implies that $\sum b_i = 0$. That is

$$\sum_{e_i^+ > 0} b_i + \sum_{e_i^- > 0} b_i + \sum_{e_i^+ = e_i^- = 0} b_i = 0,$$

which can be rewritten as

$$pP + (p-1)N + \sum_{e_i^+ = e_i^- = 0} b_i = 0. \qquad \square$$

From Proposition 7.2.2, we can derive Theorem 2.2 in Koenker (2005), which says

$$N \leq np \leq N + Z$$

$$P \leq n(1-p) \leq P + Z.$$

As a consequence, the proportion of negative residuals is approximate p and the proportion of positive residuals is approximate $1-p$. That is,

$$\frac{N}{n} \leq p \leq \frac{N+Z}{n}, \quad \text{and} \quad \frac{P}{n} \leq 1-p \leq \frac{P+Z}{n}.$$

This result is consistent with that in the case of population. As $n \to \infty$ and $Z/n \to 0$, we have that the proportion of negative residuals tends to p and the proportion of positive residuals rends to $1-p$. In general, any quantile regression line splits the whole space into two parts; the lower part has about p proportion of the observations and the upper part has about $1-p$ proportion of the observations.

7.2.2 Directional regression quantiles and halfspace depth

Laine (2001), see also Koenker (2005), has investigated the use of quantile regression methods as a means of defining directional multivariate quantiles that can be related to halfspace depth.

Definition 7.2.1 *Given a sample $\{x_1, x_2, \cdots, x_n\}$ of a random variable X in $\mathbb{R}^d$ and a vector $v \in \mathbb{B}^{d-1} = \{v : \ \|v\| \leq 1, \ v \in \mathbb{R}^d\}$, we denote the projection of x_i onto the direction $v/\|v\|$ by $x_{i,v}$ and its orthogonal complement by $x_{i,v}^{\perp}$, respectively. The directional regression quantile line on direction v is obtained by solving*

$$\min_{(\beta,\gamma) \in \mathbb{R}^{d-1} \times \mathbb{R}} \sum_{i=1}^{n} \rho_{\|v\|} \left(x_{i,v} - (x_{i,v}^{\perp})^T \beta - \gamma \right),$$

where $\rho_{\|v\|}$ is the "check" function $\rho_{\|v\|}(x) = x(\|v\| - I(x < 0))$.

Denote the projection of X onto the direction $v/\|v\|$ by X_v and its orthogonal complement by $X_v^{\perp}$, respectively. The population version of Definition 7.2.1 is to solve

$$\min_{(\beta,\gamma)\in\mathbb{R}^{d-1}\times\mathbb{R}} \sum_{i=1}^{n} \rho_{\|v\|}\left(X_v - (X_v^{\perp})^T\beta - \gamma\right).$$

If the probability distribution P of the random variable X is absolutely continuous, taking derivative on the above equation, we have the following gradient condition,

$$E\left[\|v\| - I(X_v \leq (X_v^{\perp})^T\beta + \gamma)\right]\begin{pmatrix} X_v^{\perp} \\ 1 \end{pmatrix} = 0.$$

The contribution of the "intercept" term γ requires that at a solution,

$$\|v\| = P\{X_v \leq (X_v^{\perp})^T\beta(v) + \gamma(v)\},$$

and so the mass below the hyperplane $\{X_v \leq (X_v^{\perp})^T\beta(v) + \gamma(v)\}$ is $\|v\|$. Laine (2001) defined quantile contour sets as the intersection of these lower halfspaces and pointed that the defined multivariate quantiles were related yo halfspace depth.

To study the connection between directional regression quantiles and halfspace depth, we will focus our attention on the two dimensional case. For higher dimensional case, we can obtain analogous results using the same technique. The following proposition plays an important role in verifying the relationship between directional regression quantiles and halfspace depth.

Proposition 7.2.3 *Suppose we have a line l in $\mathbb{R}^2$ with equation $y = \tan(\beta)x + \gamma$ and a data point $(x, y) \in \mathbb{R}^2$. We perform a rotation transformation about the origin with rotation angle α. Denote the transformed line by l_α with equation $y_\alpha = \tan(\beta_\alpha)x_\alpha + \gamma_\alpha$ and the transformed data point by (x_α, y_α), respectively. If $\tan(\beta)\tan(\beta_\alpha) \neq 0$, then*

$$\cos(\beta)e = \cos(\beta_\alpha)e_\alpha,$$

where e is the residual of the data point (x, y) with respect to the line l and e_α is the residual of the transformed data point (x_α, y_α) with respect to the transformed line l_α, respectively.

Proof. The relationship between data point (x, y) and the transformed data point (x_α, y_α) is

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x_\alpha \\ y_\alpha \end{pmatrix}.$$

Plugging it into the line equation $y = \tan(\beta)x + \gamma$, we have

$$y_\alpha = \tan(\alpha + \beta)x_\alpha + \cos(\beta)/\cos(\alpha + \beta)\gamma.$$

That is, $\tan(\beta_\alpha) = \tan(\alpha + \beta)$ and $\gamma_\alpha = \cos(\beta)/\cos(\alpha + \beta)\gamma$. The rotation formula between (x, y) and (x_α, y_α) says that

$$x_\alpha = \cos(\alpha)x - \sin(\alpha)y$$

$$y_\alpha = \sin(\alpha)x + \cos(\alpha)y.$$

Therefore, we have

$$\begin{aligned} e_\alpha &= y_\alpha - \tan(\alpha + \beta)x_\alpha - \cos(\beta)/\cos(\alpha + \beta)\gamma \\ &= \cos(\beta)/\cos(\alpha + \beta)\left[y - \tan(\beta)x - \gamma\right] \\ &= \cos(\beta)/\cos(\alpha + \beta)e. \end{aligned}$$

That is, $\cos(\beta)e = \cos(\beta_\alpha)e_\alpha$ holds true. □

If α is not so large, then $\cos(\beta)$ and $\cos(\alpha + \beta)$ will have the same signs. As a consequence of Proposition 7.2.3, e and e_α have the same signs. That is, if the rotation angle is small enough, then the residuals after rotation do not change signs. The result still holds true if the rotation is around a general point rather than the origin. The following theorem says that any directional regression quantile line is a boundary line of a depth contour and any boundary line of a depth contour is a directional regression quantile line.

Theorem 7.2.1 *For $0 < p < 1$, any p-th directional regression quantile line in $\mathbb{R}^2$ is a boundary line of the p-th depth contour. Conversely, any boundary line of the p-th depth contour is a p-th directional regression quantile line.*

Proof. Let l be a p-th directional regression quantile line on the direction v. The coefficients of l are obtained by solving

$$\min_{(\beta,\gamma)\in\mathbb{R}\times\mathbb{R}} \sum_{i=1}^{n} \rho_{\|v\|}\left(y_{i,v} - x_{i,v}\beta - \gamma\right), \tag{7.3}$$

where $y_{i,v}$ is the projection of (x_i, y_i) onto the direction $v/\|v\|$ and $x_{i,v}$ is its orthogonal complement, $i = 1, 2, \cdots, n$. The minimization problem (7.3) can be rewritten as a linear program problem that is a special case of linear problem (7.1). As a consequence of Proposition 7.2.2, we have

$$\frac{N}{n} \leq p \leq \frac{N+Z}{n}, \quad \text{and} \quad \frac{P}{n} \leq 1-p \leq \frac{P+Z}{n},$$

where P is the number of observations with positive residuals, N is the number of observations with negative residuals, and Z is the number of exactly fitted observations. Therefore, line l is a boundary line of a halfspace containing at least $n - \lceil np \rceil + 1$ data points. Proposition 7.1.2 implies that l is a boundary line of the p-th depth contour.

Conversely, let l be a boundary line of the p-th depth contour. Line l is a directional regression quantile line if and only if $\sum_{i=1}^{n}(p-b_i)e_{i,\alpha}^{+} = 0$ and $\sum_{i=1}^{n}(1-p+b_i)e_{i,\alpha}^{-} = 0$, where $e_{i,\alpha}$ is the i-th residual after a rotation with angle α, $p-1 \leq b_i \leq p$, and $\sum x_{i,\alpha}b_i = 0$; see Proposition 7.2.1. Similar to the proof of Proposition 7.2.2, we have: if $e_{i,\alpha}^{+} > 0$, then $b_i = p$ and if $e_{i,\alpha}^{-} > 0$, then $b_i = p-1$. Therefore, we have

$$p \sum_{e_{i,\alpha}^{+}>o} x_{i,\alpha} + (p-1) \sum_{e_{i,\alpha}^{-}>0} x_{i,\alpha} + \sum_{e_{i,\alpha}^{+}=e_{i,\alpha}^{-}=0} x_{i,\alpha}b_i = 0,$$

where $p-1 \leq b_i \leq p$. If we assume that α is small enough, we obtain, in view of Proposition 7.2.3,

$$\begin{aligned} &\cos(\alpha)\left(p \sum_{e_i^{+}>o} x_i + (p-1) \sum_{e_i^{-}>0} x_i + \sum_{e_i^{+}=e_i^{-}=0} x_i b_i\right) \\ &\quad = \sin(\alpha)\left(p \sum_{e_i^{+}>o} y_i + (p-1) \sum_{e_i^{-}>0} y_i + \sum_{e_i^{+}=e_i^{-}=0} y_i b_i\right), \end{aligned} \tag{7.4}$$

where $p - 1 \leq b_i \leq p$. According to Proposition 7.1.2, the boundary line l is the boundary of a halfspace containing at least $n - \lceil np \rceil + 1$ data points. Hence the number of nonnegative residual data points or the number of nonpositive residual data points is $n - \lceil np \rceil + 1$, which will guarantee that there exists a solution to equation (7.4). Therefore, the boundary line l is a p-th directional regression quantile line on the direction $(\cos(\alpha), \sin(\alpha))$ where α is a solution of equation (7.4). □

Proposition 7.1.1 implies that if $\lceil np_1 \rceil = \lceil np_2 \rceil$, then $D_{p_1} = D_{p_2}$. Therefore, if $\lceil np_1 \rceil = \lceil np_2 \rceil$, then any p_1-th directional regressions quantile line can be a boundary line of the p_2-th depth contour. Conversely, any boundary line of the p_1-th depth contour can be a p_2-th directional regression quantile line. Therefore, in Theorem 7.2.1, the quantile indices of depth contours and directional regression quantile lines are not necessarily equal. Note that the solutions to (7.4) are not unique. As a consequence, in Theorem 7.2.1, there exist more than one p-th directional regression quantile lines (with different directions) corresponding to any boundary line of the p-th depth contour. Nevertheless, we proved the "one-on-one" relationship between directional regression lines and boundary lines of depth contours. The multivariate quantile contours defined by Laine (2001) are equivalent to the depth contours. As an application of Theorem 7.2.1, we can compute depth contour by computing quantile regression, whose algorithm has been implemented in many existing software (e.g. `R`). To this end, the following proposition may be helpful.

Proposition 7.2.4 *Suppose that l_1, l_2 and l_3 are three directional regression quantiles adjacent to each other in counterclockwise order. If their quantile indices are the same, then $\alpha_1 \leq \alpha_2 \leq \alpha_3$ where α_1, α_2 and α_3 are their corresponding direction angles, respectively, and $-\pi < \alpha_i < \pi$ for $i = 1, 2, 3$.*

Proof. Denote the common quantile index by p. Without loss of generality, we assume that the p-th depth contour D_p is above l_i, $i = 1, 2, 3$ and $\alpha_2 = 0$. The rotation angle $\alpha_2 = 0$ means that l_2 is the usual p-th regression quantile without any rotation. To prove $\alpha_1 \leq \alpha_2 \leq \alpha_3$, we only need to show that $\alpha_1 \leq 0$ and $\alpha_2 \geq 0$. Let $E_i = \sum_{j=1}^{n} pe_{ij}^{+} + (1-p)\sum_{i=1}^{n} e_{ij}^{-}$ denote the sum of weighted residuals with respect to l_i where e_{ij} is the residual of the j-th observation with respect to l_i, $e_{ij}^{+} = e_{ij} \vee 0$ and $e_{ij}^{-} = (-e_{ij}) \vee 0$. Obviously we have $E_1 \geq E_2$ and $E_3 \geq E_2$ because that l_2 is the p-th quantile regression line. Suppose that the slope angle of l_1, l_2 and l_3 are β_1, β_2 and β_3, respectively. Then $\tan\beta_1 < \tan\beta_2 < \tan\beta_3$. Let E_i^{α}, $i = 1, 2, 3$, be the residuals after a rotation with angle α. After a rotation of angle α, $E_i^{\alpha} = E_i \cos\beta_i / \cos(\beta_i + \alpha)$ in view of Proposition 7.2.3 and the ratio of the sum of weighted residuals E_3^{α} to E_2^{α} is

$$\frac{E_3^{\alpha}}{E_2^{\alpha}} = \frac{E_3 \cos\beta_3 \cos(\beta_2\alpha)}{E_2 \cos\beta_2 \cos(\beta_3 + \alpha)} = \frac{1 - \tan\beta_2 \tan\alpha}{1 - \tan\beta_3 \tan\alpha} \frac{E_3}{E_2}.$$

If α is the direction angle of directional regression quantile line l_3, we will have $E_3^{\alpha}/E_2^{\alpha} \leq 1$ as $E_3/E_2 \geq 1$. Therefore, we reach $(1-\tan\beta_2\tan\alpha)/(1-\tan\beta_3\tan\alpha) \leq 1$. The slope relationship of $\tan\beta_2 < \tan\beta_3$ implies that α has to be no less than zero. That is, $\alpha_3 \geq 0$. Similarly, we have $\alpha_2 \leq 0$. □

Chapter 8

Quantiles in general statistical models

8.1 Introduction

We propose a general definition of the quantiles based on the halfspace depth. The definition was inspired by the depth-based regression quantile proposed by Roger Koenker, documented by Rousseeuw and Hubert (1999) and further investigated by Adrover, Maronna and Yohai (2004). The depth-based definition of regression quantile is an alternative to the L_1-based one known from Koenker and Bassett (1978). Our definition can be viewed as its generalization that is capable of defining quantiles in various data-analytic situations.

In well-behaved regression models, both Koenker's depth-based quantile and the L_1 regression quantile estimate the same quantity (for equal p). The Koenker's depth-based proposal was further developed by Adrover, Maronna and Yohai (2004) as one of the possibilities to estimate quantiles in a robust way—more precisely, as a leverage-robust alternative to the standard L_1 procedure. Our objective here is not that much robustness as a unified definition of quantiles, applicable in various statistical models and data-analytic situations. Despite our population orientation, it is the sample setting that provides a more understandable formal language—more than that of abstract measures. Therefore, we start our explanation from empirical distributions.

In Section 8.2, we advance Koenker's idea to a new concept of depth quantile in the general framework of the theory of tangent depth (Mizera, 2002). We show our quantiles are affine equivariant in the multivariate location setting in Theorem 8.3.1. We study the depth quantiles of elliptically symmetric distributions and give their half closed form in Theorem 8.3.2. We investigate the relationship to the special case of univariate location and present the result in Theorem 8.4.1. In regression with intercept, we clarify the relationship to the original Koenker's depth-based regression quantile in Theorem 8.5.2. We point out certain differences arising in regressions without intercept and gives an example showing that the L_1-based quantile regression by Koenker and Bassett (1978) fails, while our quantile regression provides reasonable results.

8.2 Depth quantiles: definition and discussion

Given a particular data-analytic situation, we start by specifying a collection of criterial functions $F_1, F_2, \cdots, F_n$ defined on a space Θ parameterizing the fits; in what follows, these criterial functions will be considered fixed. Given data points z_i, $i = 1, 2, \cdots, n$, we define the depth of ϑ to be

$$d(\vartheta) = \inf_{u \neq 0} P_n(\{z_i : \nabla F_i(\vartheta) \in H_u\}) = \inf_{u \neq 0} P_n(\{z_i : u^T \nabla F_i(\vartheta) \geq 0\}), \tag{8.1}$$

where P_n is the empirical distribution and $H_u = \{x : u^T x \geq 0\}$ denotes the closed halfspace that contains the origin on its boundary, with the vector u pointing inside the halfspace and orthogonal to the boundary. See Mizera (2002), Mizera and Müller (2004), and the examples below for how the criterial functions F_i correspond with data points z_i, and how the structure of both reflects the data-analytic situation under consideration.

The depth-based analog of median is customarily called maximum depth estimator in the literature; see Donoho and Gasko (1992), Rousseeuw and Hubert (1999), Mizera

(2002), Mizera and Müller (2004), and others. The name indicates that its value is any $\vartheta \in \Theta$ maximizing the depth for the fixed data set. Similar to the univariate median, non-uniqueness complications may arise; however, we believe that they are of a minor nature and will not be dealt with here.

The definition of depth quantiles proceeds in a similar vein, via the maximization of a "quantile depth", a depth whose definition is modified with respect to the index u and direction w.

Definition 8.2.1 *For any given $p \in (0,1)$ and any $w \neq 0$, a* depth quantile $\vartheta_{p,w}$ *is defined as ϑ maximizing the* quantile depth

$$d_{p,w}(\vartheta) = \inf_{u \neq 0} \left(2(1-p)\mathcal{F}(H_u \cap H_w) + 2p\mathcal{F}(H_u \cap H_{-w}) - \mathcal{F}(H_u \cap H_w \cap H_{-w})\right), \tag{8.2}$$

where $\mathcal{F}(E)$ is a shorthand for $P_n(\{z_i : \nabla F_i(\vartheta) \in E\})$.

Note that the depth quantile depends only on the orientation of w: $\vartheta_{p,cw} = \vartheta_{p,w}$ whenever $c > 0$. When $p = 1/2$, then (8.2) reduces to (8.1), that is, $d_{1/2,w}(\vartheta) = d(\vartheta)$; hence, $\vartheta_{1/2,w} = \vartheta_{1/2}$, the deepest fit, for every w. The definition following the original formulation given by Rousseeuw and Hubert (1999) would use

$$2 \inf_{u \neq 0} \left((1-p)\mathcal{F}(H_u \cap H_w) + p\mathcal{F}(H_u \cap H_{-w})\right) \tag{8.3}$$

instead of (8.2). However, such a definition would not reduce to the maximum depth estimator for $p = 1/2$; also, it might yield different results for different w, even if $p = 1/2$. Another possibility consisting in replacing (8.2) with

$$2 \inf_{u \neq 0} \left((1-p)\mathcal{F}(H_u \cap H_w) + p\mathcal{F}(H_u \backslash H_w)\right) \tag{8.4}$$

would not suffer from this drawback; however, it would not satisfy the *antipodal identity* $\vartheta_{p,w} = \vartheta_{1-p,1-w}$, which the symmetric definition satisfies for every p and w. We will call the quantiles indexed by (p,w) and $(1-p,-w)$ *antipodal*.

Once the antipodal identity holds true, for instance, for definitions based either on (8.2) or (8.3), we can identify antipodal quantiles and subsequently limit p to the interval $[1/2, 1)$. The independence on w for $p = 1/2$ allows further for identifying all quantiles with $p = 1/2$, so that, after rescaling, we can index quantiles by the elements of the unit disk $\{w : \|w\| \leq 1\}$. However, the antipodal identity suggests also an alternative indexing, which also identifies antipodal quantiles, but then leaves p in the interval $(0, 1)$ and views the index w as an element of an appropriate projective plane (for instance, $\{w : u^T w \geq 0, u \neq 0\}$).

8.3 Location: multivariate

In this instance, the criterial functions $F_1, F_2, \cdots, F_n$ are in mutual correspondence with the data points $z_1, z_2, \cdots, z_n$. A possible choice is $F_i = \frac{1}{2}\|z_i - \vartheta\|^2$, with gradients $\nabla F_i(\vartheta) = (\vartheta - z_i)$. Consequently, $\mathcal{F}(H_u) = P_n(\{i : \vartheta - z_i \in H_u\}) = P_n(\{i : u^T(\vartheta - z_i) \geq 0\})$. The quantiles are dependent on the data points; to make this dependence more explicit in the notation, we may write $\vartheta_{p,w} = \vartheta_{p,w}(z_1, z_2, \cdots, z_n)$. For more detailed exposition, see Rousseeuw and Hubert (1999) or Mizera (2002). Note that both z_i and ϑ are elements of $\mathbb{R}^d$.

Theorem 8.3.1 *For depth quantiles in multivariate location,*

$$\vartheta_{p,\tilde{w}}(Az_1 + b, Az_2 + b, \cdots, Az_n + b) = A\vartheta_{p,w}(z_1, z_2, \cdots, z_n) + b, \qquad (8.5)$$

for any $p \in (0, 1)$, $w \in \mathbb{R}^p$, any $d \times d$ nonsingular matrix A, any vector b in $\in \mathbb{R}^p$, and every $z_1, z_2, \cdots, z_n$, where $\tilde{w} = (A^T)^{-1}w\|w\|/\|(A^T)^{-1}w\|$.

Proof. A straightforward verification. Let $\tilde{z}_i = Az_i + b$, $i = 1, 2, \cdots, n$, and $\tilde{\vartheta} = A\vartheta + b$. We have $u^T(\tilde{z}_i - \tilde{\vartheta}) = u^T(Az_i - A\vartheta) = (A^T u)^T(z_i - \vartheta)$; therefore, if $\tilde{w} = (A^T)^{-1}w\|w\|/\|(A^T)^{-1}w\|$, then the sign of $\tilde{w}^T(\tilde{z}_i - \tilde{\vartheta})$ is the same as that of

$w^T(z_i - \vartheta)$. Since there is one-one correspondence between u and $A^T u$, it follows that $\tilde{\vartheta}$ maximizes (8.2) with respect to the transformed data points $\tilde{z}_i$ if and only if ϑ maximizes (8.2) with respect to the original data points z_i, $i = 1, 2, \cdots, n$. □

We argue that this transformation rule expresses the affine equivariance for the depth quantiles. The quantile index p always remains invariant. If $p = 1/2$, then w is inessential and formula (8.5) expresses the affine equivariance of the Tukey median. In other cases, the direction w may change with the transformation. Indeed: if, for instance, A corresponds to a rotation, then $(A^T)^{-1} = A$ and the quantile of rotated datapoints is, very naturally, the rotated quantile in the rotated direction.

Interestingly, the transformation rule for affine-equivariant quantiles of Chakraborty (2001) has for non-orthogonal A a form different from (8.5); it takes the transformed direction w to be directly that of Aw. Let $\eta_{p,w}$ denote the p-th quantile with direction w defined by Chakraborty (2001), then its affine equivariance says

$$\eta_{p,\tilde{w}}(Az_1 + b, Az_2 + b, \cdots, Az_n + b) = A\eta_{p,w}(z_1, z_2, \cdots, z_n) + b,$$

where $\tilde{w} = Aw\|w\|/\|Aw\|$. Nevertheless, both rules are valid expressions of affine equivariance; their differences lie in the style of indexing of quantiles, which can be best explained on multivariate normal distributions. Before that, we need to formulate the measure-theoretic definition of depth quantiles in multivariate location. The symbol $H_{\vartheta,u}$ will denote the closed halfspace $\{x : u^T(x-\vartheta) \geq 0\}$ containing ϑ on its boundary, with vector u orthogonal to it and pointing inside the halfspace.

Definition 8.3.1 *For any probability measure P on $\mathbb{R}^d$, we define the* location depth quantile $\vartheta_{p,w} = \vartheta_{p,w}(P)$*, for given $p \in (0,1)$ and any $w \neq 0$ as any ϑ maximizing*

$$\begin{aligned} d_{p,w}(\vartheta) = \inf_{u \neq 0} (2(1-p)P(H_{\vartheta,u} \cap H_{\vartheta,w}) + 2pP(H_{\vartheta,u} \cap H_{\vartheta,-w}) \\ - P(H_{\vartheta,u} \cap H_{\vartheta,w} \cap H_{\vartheta,-w})). \end{aligned} \tag{8.6}$$

The transformation rule expressed by Theorem 8.3.1 remains valid, formalizing the relationship between $\vartheta_{p,w}(P)$ and $\vartheta_{p,w}(PT^{-1})$, where $T(x) = Ax + b$. The following proposition characterizes the location depth quantiles of certain symmetric distributions. For simplicity, the proposition is formulated for nonsingular P, although it holds in general. In singular cases, we have only to consider restriction of P to its supporting affine subspace and consider its density there; this implies the natural restriction of interesting indexing directions (the quantiles in "singular" directions coincide for all p).

Proposition 8.3.1 *For any nonsingular probability measure P in $\mathbb{R}^2$, if there exists a line l such that P is symmetric about it, then there must be a $\vartheta_{p,w}(p)$ which lies on the line l, for any $p \in (0,1)$ and w collinear with l.*

Proof. Suppose that A is an arbitrary point on l. We denote the line perpendicular to l and passes through A by m. For $p \in (0,1)$ and w collinear with l, we have the quantile depth of A is no less than any other point lying on m. For any point not on l, there exists a point on l with greater quantile depth that that point. For any point B on m, denoting $\{x : u^T(x - B) \geq 0\}$ by $H_{B,u}$, it has quantile depth $d_{p,w}(B)$ with the following form,

$$\begin{aligned}
&\inf_{u \neq 0} \{2pP(H_{B,u} \cap H_{B,w}) + 2(1-p)P(H_{B,u} \cap H_{B,-w}) - P(H_{B,u} \cap H_{B,w} \cap H_{B,-w})\} \\
=&\inf_{u \neq 0} \{2pP(H_{B,u} \cap H_{B,w}) + 2(1-p)P(H_{B,u} \cap H_{B,-w})\} \\
=&c \inf_{u \neq 0} \{P^*(H_{B,u})\},
\end{aligned}$$

where $c = 2pP(H_{A,w}) + 2(1-p)P(H_{A,w})$ is a constant factor for fixed point A and P^* is defined as,

$$P^*(E) = \frac{2pP(E \cap H_{A,w}) + 2(1-p)P(E \cap H_{A,-w})}{2pP(H_{A,w}) + 2(1-p)P(H_{A,w})}.$$

It is trivial that P^* is an probability measure. Suppose that $d_{p,w}(B)$ is obtained at point B when $u = v$. There are two possibilities: either $w^T v \geq 0$ or $w^T v \leq 0$. For the first

case, $w^T v \geq 0$, if $B \in H_{A,v}$ (i.e. B is above the line l), we can obtain that $H_{B,v} \subseteq H_{A,v}$. Then from the definition of P^*, we know that $P^*(H_{B,v}) \leq P^*(H_{A,v})$, which means

$$d_{p,w}(A) = cP^*(H_{A,v}) \geq cP^*(H_{B,v}) \geq d_{p,w}(B).$$

If $B \notin H_{A,v}$, we can find its counterpart point symmetric about l, say B'. From the discussion above, we know $d_{p,w}(A) \geq d_{p,w}(B') = d_{p,w}(B)$, since $B' \in H_{A,v}$. Wherever B, we can obtain that $d_{p,w}(A) \geq d_{p,w}(B)$ as long as B is lying on m. Similarly, for the second case, $w^T v \leq 0$, $d_{p,w}(A) \geq d_{p,w}(B)$ holds as well. □

Consequently, if P is a rotationally symmetric distribution, then all the points lying on the same cycle with a symmetric center as cycle center have equal quantile depth and their corresponding index directions are perpendicular to the cycle. This property holds not only for $\mathbb{R}^2$ but also for any $\mathbb{R}^d$. Specifically, we have the following theorem for elliptically symmetric distributions.

Theorem 8.3.2 *Suppose that P is a nonsingular elliptically symmetric distribution, with the expectation μ and variance-covariance matrix Σ. For fixed p , all location depth quantiles $\vartheta_{p,w}$ lie on an ellipsoid surface of constancy of the density of P:*

$$\vartheta_{p,w} = \mu + k_p \frac{\Sigma w}{\|\Sigma^{1/2} w\|}, \tag{8.7}$$

where the unique k_p depends only on p. The position of the quantile on the ellipsoid surface is characterized by the direction u of the halfspace $H_{\vartheta,u}$ minimizing (8.6) for $\vartheta = \vartheta_{p,w}$; it coincides with that of w.

Proof. The first step is to take the standardizing transformation $T(x) = \Sigma^{-1/2}(x - \mu)$. The transformed distribution, $\tilde{P}$, is rotationally symmetric with mean 0 and unit diagonal variance-covariance matrix I. If $\vartheta_{p,w}(\tilde{P}) = k_p w/\|w\|$, in view of Theorem 8.3.1, we have $\vartheta_{p,\Sigma^{-1/2}w}(P) = \mu + k_p \Sigma^{1/2} w/\|w\|$. The formula (8.7) then follows by the change of variable w. □

The affine-equivariant multivariate location quantiles of Chakraborty (2001) lie also, for fixed p, on an ellipsoid surface of constant density (possibly different from ours for the same p). However, his quantile always lies on its indexing direction: $w/\|w\| = \vartheta_{p,w}/\|\vartheta_{p,w}\|$. As a result of Theorem 8.3.2, we have only to determine the quantiles of $\tilde{P}$, rotationally symmetric distribution with mean 0 and unit diagonal variance-covariance matrix I. In view of Proposition $8.3.1$ and the rotational symmetry of $\tilde{P}$, we have only to consider $w = (1, 0, \cdots, 0)$. Unfortunately, we are not able to obtain an explicit formula for k_p, even for multivariate standard normal distribution. It would be ideal if k_p would be equal to $\Phi^{-1}((1+p)/2)$, where Φ denotes the cumulative distribution function of the standard normal distribution. In such a case the depth quantile would correspond to the marginal quantile in the direction given by w. This is not the case. Simulation results show that the numerical value of k_p for the two-dimensional standard normal distribution is always a little bit greater than the ideal univariate p-th quantile, $\Phi^{-1}((1+p)/2)$, for the standard normal distribution.

8.4 Location: univariate

The special case of the location setting that deserves particular attention is the one-dimensional one, with $p = 1$. Obviously, we are interested in how our general definition is compatible with the classical univariate one. Recall that for a probability measure P on $\mathbb{R}$, the classical definition defines the quantile corresponding to $p \in (0, 1)$ to be any $Q(p)$ such that simultaneously $P((-\infty, Q(p)]) \geq p$ and $P([Q(p), \infty)) \leq 1 - p$. In the one-dimensional case, there are only two directions: $w = \pm 1$, and the antipodal identity further reduces the essential number of choices to one: $w = 1$ or $w = -1$.

Theorem 8.4.1 *Suppose that P is a probability measure P on $\mathbb{R}$. If $P(\{x\}) = 0$ for every $x \in \mathbb{R}$, then $\vartheta_{p,-1} = Q(p)$ for every $p \in (0, 1)$.*

Proof. Definition 8.3.1 of $d_{p,-1}(\vartheta)$ says $\vartheta_{p,-1}$ is the solution of

$$\min\left\{2(1-p)P(H_{\vartheta,1}\cap H_{\vartheta,-1})+2pP(H_{\vartheta,1}), 2(1-p)P(H_{\vartheta,1})+2pP(H_{\vartheta,-1}\cap H_{\vartheta,1})\right\},$$

which is equivalent to, in the view of $P(H_{\vartheta,-1}\cap H_{\vartheta,1})=P(\{\vartheta\})=0$,

$$\begin{aligned}&2\min\left\{pP([\vartheta,\infty)),(1-p)P((-\infty,\vartheta])\right\}\\&\quad=2\min\left\{p(1-P((-\infty,\vartheta])),(1-p)P((-\infty,\vartheta])\right\}.\end{aligned}$$

The latter is to minimize a decreasing function and an increasing function. Its minimum value is obtained when the two functions are equal, that is,

$$p(1-P((-\infty,\vartheta_{p,-1}]))=(1-p)P((-\infty,\vartheta_{p,-1}]),$$

which yields $P((-\infty,\vartheta_{p,-1}])=p$ and consequently $P([\vartheta_{p,-1},\infty))=1-p$. □

Unfortunately, Theorem $8.4.1$ may not hold true for probability measures with a discrete component. As an example, take P to be the empirical distribution of five data points located at $1,2,3,4,5$. While $Q(0.25)=Q(0.30)=Q(0.35)=2$, the direct computation shows that although $\vartheta_{0.25,-1}=\vartheta_{0.30,-1}=2$, the location depth quantile $\vartheta_{0.35,-1}=3$. This anomaly would be an opportunity to reconsider once again the exact form of formula (8.2) appearing in Definition $8.2.1$; however, neither (8.3) nor (8.4) would give a different result. The definition (8.3) with open halfspaces in place of H_u would be able to pick the correct quantity 2 for the quantile, but would add to it all the elements of the interval $[2,3]$ as well.

8.5 Linear regression with intercept

In a linear regression model, every data point consists of a response part and a covariate part: $z_i=(y_i,x_i^T)^T\in\mathbb{R}\times\mathbb{R}^d$, $i=1,2,\cdots,n$. A possible choice for criterial functions is to set $F_i(\vartheta)=\frac{1}{2}(y_i-x_i^T\vartheta)^2$; their gradients are $\nabla F_i(\vartheta)=-(y_i-x_i^T\vartheta)x_i=-e_ix_i$,

where e_i is the residual of i-th observation. The notation reflects the dependence on the data points. For more background, see Rousseeuw and Hubert (1999) or Mizera (2002).

Theorem 8.5.1 *Given data points $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, $i = 1, 2, \cdots, n$, let $p \in (0, 1)$ and $w \in \mathbb{R}^d$. For any depth quantile $\vartheta_{p,w}(z_1, z_2, \ldots, z_n)$ in a regression model, the following properties hold. For every $v \in \mathbb{R}^d$, if $\tilde{z}_i = (\tilde{y}_i, x_i^T)^T$ and $\tilde{y}_i = y_i + v^T x_i$, then*

$$\vartheta_{p,w}(\tilde{z_1}, \tilde{z_2}, \ldots, \tilde{z_n}) = \vartheta_{p,w}(z_1, z_2, \ldots, z_n) + v. \tag{8.8}$$

For every $c > 0$, if $\tilde{z}_i = (\tilde{y}_i, x_i^T)^T$ and $\tilde{y}_i = cy_i$, then

$$\vartheta_{p,w}(\tilde{z_1}, \tilde{z_2}, \ldots, \tilde{z_n}) = c\vartheta_{p,w}(z_1, z_2, \cdots, z_n). \tag{8.9}$$

For every nonsingular $d \times d$ matrix A, if $\tilde{z}_i = (y_i, \tilde{x_i}^T)^T$, $\tilde{x_i} = Ax_i$, and $\tilde{w} = (A^T)^{-1}w$, then

$$\vartheta_{p,w}(\tilde{z_1}, \tilde{z_2}, \ldots, \tilde{z_n}) = (A^T)^{-1}\vartheta_{p,w}(z_1, z_2, \ldots, z_n). \tag{8.10}$$

Proof. If $\tilde{y_i} = y_i + v^T x_i$ for $i = 1, 2, \cdots, n$, then the transformed gradients are $\nabla F_i(\tilde{\vartheta}) = -x_i(\tilde{y_i} - x_i^T\tilde{\vartheta}) = -x_i(y_i - x_i^T(\tilde{\vartheta} - v))$. Data point $\tilde{\vartheta}$ maximizes (8.2) with respect to the transformed data points $\{\tilde{z_1}, \tilde{z_2}, \cdots, \tilde{z_n}\}$ if and only if $\tilde{\vartheta} - v$ maximizes (8.2) with respect to the original data points $\{z_1, z_2, \cdots, z_n\}$. Therefore, (8.8) holds true.

If $\tilde{y_i} = cy_i$ for $i = 1, 2, \cdots, n$, then the transformed gradients are $\nabla F_i(\tilde{\vartheta}) = -x_i(\tilde{y_i} - x_i^T\tilde{\vartheta}) = -cx_i(y_i - x_i^T\tilde{\vartheta}/c)$; then $\tilde{\vartheta}$ maximizes (8.2) if and only if $\tilde{\vartheta}/c$ maximizes (8.2). Therefore, (8.9) holds true.

If $\tilde{x_i} = Ax_i$ for $i = 1, 2, \cdots, n$, then transformed residuals are $\tilde{e_i} = y_i - \tilde{x_i}^T\tilde{\vartheta} = y_i - (Ax_i)^T\tilde{\vartheta} = y_i - x_i^T(A^T\tilde{\vartheta}) = e_i$ if and only if $A^T\tilde{\vartheta} = \vartheta$. If it does, we can obtain that $\tilde{u}^T\nabla F_i(\tilde{\vartheta}) = -\tilde{u}^T\tilde{x_i}\tilde{e_i} = -(A^T\tilde{u})^T\nabla F_i(\vartheta)$ and $\tilde{w}^T\nabla F_i(\tilde{\vartheta}) = -\tilde{w}^T\tilde{x_i}\tilde{e_i} =$

$(A^T\tilde{w})^T\nabla F_i(\vartheta)$. As $\tilde{u}$ can be any nonzero vector in $\mathbb{R}^d$ and A is a nonsingular matrix, u can also be any nonzero vector in $\mathbb{R}^d$. That is, $\tilde{u} \neq 0 \Leftrightarrow u \neq 0$. Hence, if we let $\tilde{w} = (A^T)^{-1}w$, we obtain that $\tilde{\vartheta}$ maximize (8.2) with respect to the transformed data points if and only if $\vartheta = A^T\tilde{\vartheta}$ maximizes (8.2) with respect to the original data points. Therefore, (8.10) holds for every nonsingular $d \times d$ matrix A. □

The transformation rules (8.8) and (8.9) have the same forms as those establishing regression and scale equivariance of regression estimators. Hence, the only question is how much can be (8.10) viewed as an expression of affine equivariance. For $p = 1/2$, direction w is again inessential and (8.10) means usual affine equivariance of the deepest regression. If A is orthogonal, then $(A^T)^{-1} = A$, and the transformation rule is natural: the transformed quantile with the same p and direction Aw is $A\vartheta_{p,w}$.

Suppose that the regression contains intercept: $x_i = (t_i^T, 1)^T$. Any meaningful transformation of regressors in this case leaves intercept fixed and any transformation matrix A thus has the form

$$A = \begin{pmatrix} B & u \\ 0 & 1 \end{pmatrix}.$$

In particular, when B is an identity matrix I, then the action of A shifts all regressors by u. The transformation rule (8.10) then gives that $w = A^T\tilde{w}$, Writing $\tilde{w}^T = (\tilde{w}_1^T, \tilde{w}_0)$ and $w^T = (w_1^T, w_0)$, we obtain that $w_1 = \tilde{w}_1$ and $w_0 = \tilde{w}_1^T v + \tilde{w}_0$. The equality $w = \tilde{w}$ holds true for all v if and only if $w_1 = 0$. In such a case, the quantiles transform according to (8.10), leaving w the same. Since all w with $w_1 = 0$ represent the same projective direction, we have a special case of depth quantile in regression which is independent on w: the intercept quantile. More precisely, we have two intercept quantiles, depending on the sign of the last element of w; but they are closely related via the correspondence between p and $1 - p$. We will show now that modulo this difference, the intercept quantile is precisely Koenker's depth-based regression quantile.

In the notation of Rousseeuw and Hubert (1999), given observations $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, $i = 1, 2, \cdots, n$, any affine hyperplane V in the regressor space splits the observations into two sets, denoted by $L(V)$ and $R(V)$. For $p \in (0, 1)$, Rousseeuw and Hubert (1999) defined the regression depth quantile $\vartheta^p = \vartheta^p(z_1, z_2, \cdots, z_n)$ as the hyperplane maximizing

$$2 \inf_V \left(pP_n(L^+(V)) + (1-p)P_n(R^-(V)) \right), \tag{8.11}$$

where $L^+(V) = \{z_i : x_i \in L(V), e_i \geq 0\}$, $R^-(V) = \{z_i : x_i \in R(V), e_i \leq 0\}$, $e_i = y_i - x_i^T \vartheta$ is the i-th residual and P_n is the empirical probability distribution.

Theorem 8.5.2 *Given observations $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, $i = 1, 2, \cdots, n$, if $x_i = (t_i^T, 1)^T$ and $w = (0, 0, \cdots, 0, -1)_d^T$, then*

$$\vartheta^p(z_1, z_2, \ldots, z_n) = \vartheta_{p,w}(z_1, z_2, \ldots, z_n)$$

for every $z_1, z_2, \ldots, z_n$ and $p \in (0, 1)$, where $\vartheta^p(z_1, z_2, \ldots, z_n)$ is the p-th regression depth and $\vartheta_{p,w}(z_1, z_2, \ldots, z_n)$ is the p-th depth quantile using 8.4.

Proof. For any hyperplane V, there exists a direction $u \neq 0$ such that $L(V) = \{x_i : x_i \in H_{-u}\}$, and $R(V) = \{x_i : x_i \in H_u\}$. Let $w = (0, 0, \cdots, 0, -1)_d^T$. Since $-w^T x_i = 1$, we have $w^T \nabla F_i(\vartheta) = y_i - x_i^T \vartheta = e_i$ and $u^T \nabla F_i(\vartheta) = u^T(-e_i x_i) = -e_i u^T x_i$. Hence we can rewrite $L^+(V)$ as the follow,

$$\begin{aligned} L^+(V) &= \{z_i : x_i \in H_{-u}, e_i \geq 0\} = \{z_i : -e_i u^T x_i \geq 0, e_i \geq 0\} \\ &= \{z_i : u^T \nabla F_i(\vartheta) \geq 0, w^T \nabla F_i(\vartheta) \geq 0\} \\ &= \{z_i : \nabla F_i(\vartheta) \in H_u \cap H_w\}. \end{aligned}$$

Similarly, we have

$$R^-(V) = \{z_i : \nabla F_i(\vartheta) \in H_u \cap H_{-w}\}.$$

Therefore, $\vartheta^p(z_1, z_2, \ldots, z_n)$ maximizing (8.11) is equivalent to $\vartheta^p(z_1, z_2, \ldots, z_n)$ maximizing (8.3), that is, $\vartheta^\tau(z_1, z_2, \ldots, z_n) = \vartheta_w^\tau(z_1, z_2, \ldots, z_n)$. □

If $\vartheta_{p,w}(z_1, z_2, \ldots, z_n)$ is obtained by using (8.2), then there may be a slight difference between $\vartheta_{p,w}(z_1, z_2, \ldots, z_n)$ and $\vartheta^p(z_1, z_2, \ldots, z_n)$. This difference tends to vanish withe increasing sample size.

8.6 Linear regression without intercept

In a linear regression model without intercept, suppose that we have data points $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, $i = 1, 2, \cdots, n$. Let the criterial functions to be $F_i(\vartheta) = \frac{1}{2}(y_i - x_i^T\vartheta)^2$; their gradients are $\nabla F_i(\vartheta) = -(y_i - x_i^T\vartheta)x_i$. We denote x_i by $(x_{i1}, x_{i2}, \cdots, x_{id})^T$ and ϑ by $(\vartheta_1, \vartheta_2, \cdots, \vartheta_d)^T$, respectively. If $x_{id} \neq 0$, for $i = 1, 2, \cdots, n$, then the gradients of the criterial functions can be rewritten as

$$\begin{aligned}\nabla F_i(\vartheta) &= -(y_i/x_{id} - (x_i/x_{id})^T\vartheta)x_i/x_{id}(x_{id})^2 \\ &= -(\tilde{y}_i - \tilde{x}_i^T\vartheta)\tilde{x}_i(x_{id})^2,\end{aligned}$$

where $\tilde{y}_i = y_i/x_{id}$ and $\tilde{x}_i = x_i/x_{id}$, $i = 1, 2, \cdots, n$. Therefore, given $p \in (0, 1)$ and $w \neq 0$, ϑ maximizing

$$\inf_{u \neq 0} \left(2(1-p)\mathcal{F}(H_u \cap H_w) + 2p\mathcal{F}(H_u \cap H_{-w}) - \mathcal{F}(H_u \cap H_w \cap H_{-w})\right)$$

is equivalent to ϑ maximizing

$$\inf_{u \neq 0} \left(2(1-p)\tilde{\mathcal{F}}(H_u \cap H_w) + 2p\tilde{\mathcal{F}}(H_u \cap H_{-w}) - \tilde{\mathcal{F}}(H_u \cap H_w \cap H_{-w})\right),$$

where $\mathcal{F}(E)$ is a shorthand for $P_n(\{z_i : \nabla F_i(\vartheta) \in E\})$ and $\tilde{\mathcal{F}}(E)$ is a shorthand for $P_n(\{z_i : \nabla \tilde{F}_i(\vartheta) = -(\tilde{y}_i - \tilde{x}_i^T\vartheta)\tilde{x}_i \in E\})$. Note that the first component of $\tilde{x}_i$ is 1 for all $i = 1, 2, \cdots, n$. In other words, given $p \in (0, 1)$ and $w \neq 0$, the p-th depth quantile in a linear regression model without intercept using data points $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, $i = 1, 2, \cdots, n$, is the same as the p-th depth quantile in a linear regression model using data points $\tilde{z}_i = (\tilde{y}_i, \tilde{x}_i^T)^T \in \mathbb{R} \times \mathbb{R}^d$, where $\tilde{y}_i = y_i/x_{id}$ and $\tilde{x}_i = x_i/x_{id}$, $i = 1, 2, \cdots, n$.

The latter actually is a linear regression model with intercept. In the case of $d = 1$, we will have the following proposition.

Theorem 8.6.1 *Given data points $z_i = (y_i, x_i^T)^T \in \mathbb{R} \times \mathbb{R}$, $i = 1, 2, \cdots, n$, $p \in (0, 1)$ and $w = 1$, if $x_i \neq 0$ for all $i = 1, 2, \cdots, n$, then the p-th depth quantile $\vartheta_{p,w}$ in the regression model without intercept $y \sim x$ satisfies*

$$|Q^{-1}(\vartheta_{p,w}; y_1/x_1, y_2/x_2, \cdots, y_n/x_n) - p| \leq 1/n.$$

Proof. The p-th depth quantile $\vartheta_{p,w}$ maximizes

$$\inf_{u \neq 0} \left(2(1-p)\mathcal{F}(H_u \cap H_w) + 2p\mathcal{F}(H_u \cap H_{-w}) - \mathcal{F}(H_u \cap H_w \cap H_{-w})\right),$$

where $\mathcal{F}(E)$ is a shorthand for $P_n(\{z_i : \nabla F_i(\vartheta) = -(y_i - x_i\vartheta)x_i \in E\})$. As $\nabla F_i(\vartheta) = -(y_i - x_i\vartheta)x_i = -(y_i/x_i - \vartheta)x_i^2$, $w = 1$, and $u = 1$ or $u = -1$, we have $\vartheta_{p,w}$ maximizes

$$\min\{2(1-p)P_n(\{z_i : y_i/x_i < \vartheta\}),\, 2pP_n(\{z_i : y_i/x_i > \vartheta\})\} + P_n(\{z_i, y_i/x_i = \vartheta\}).$$

Therefore,

$$|Q^{-1}(\vartheta_{p,w}; y_1/x_1, y_2/x_2, \cdots, y_n/x_n) - p| \leq 1/n. \qquad \square$$

Theorem 8.6.1 says: If the covariate is univariate, then the p-th depth quantile line without intercept is the line whose slope is approximately p-th quantile among all the slopes $\{y_1/x_1, y_2/x_2, \cdots, y_n/x_n\}$. This is a more reasonable result compared to that of L_1-based quantile regression by Koenker and Bassett (1978). In some case, even L_1-based quantile regression fails, our depth quantile regression is still capable of yielding desired regression lines, for example, in the case that univariate covariate has both negative and positive values. For simplicity, we suppose that we have 10 data points $\{(\pm 1, \pm 1.6), (\pm 1, \pm .8), (\pm 1, 0)\}$; see Figure 8.1. If the slope of a line passing through

the origin is in the interval $[-1.6, 1.6]$, then $\sum_{e_i^+>0} e_i = \sum_{e_i^-<0} e_i = 3.2$, which means $(1-p)\sum_{e_i^+>0} e_i + p\sum_{e_i^-<0} e_i$ is minimized by any of these lines for any $0 < p < 1$. Therefore, any line passing through the origin with slope in the interval $[-1.6, 1.6]$ can be a p-th quantile regression line; in particular, any of these lines is a median L_1 regression line (see Figure 8.1). On the other hand, our definition of quantiles based on depth gives a well-defined line. For any $0 < p < 1$, it corresponds to a line whose slope is the approximate p-th quantile of $\{\pm 1.6, \pm 1.6, \pm .8, \pm .8, 0, 0\}$. In particular, Figure 8.1 shows the median quantile line, the line with median slope, corresponding to the line what is known as the Theil-Sen estimator in simple regression (Rousseeuw and Leroy, 1987).

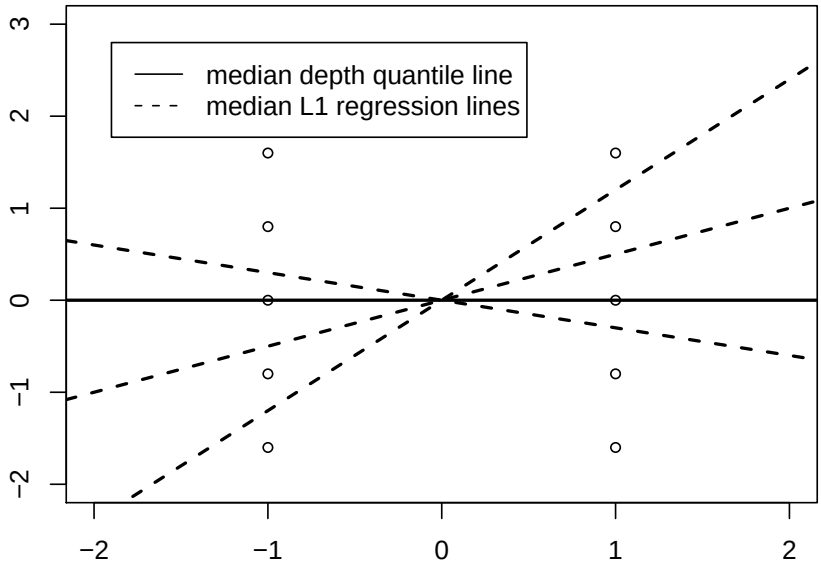

Figure 8.1: *Median L_1 regression line can be any line passing through the origin with slope on the interval $[-1.6, 1.6]$. Median depth quantile line gives the line corresponding to the Theil-Sen estimator.*

Chapter 9

Conclusions

In this thesis, we proposed multivariate quantile regression via a directional approach. Directional quantile envelopes, the intersection of supporting halfspaces determined by directional quantiles, are essentially halfspace depth contours. Directional quantile envelopes share many common merits with halfspace depth contours and also allow the addition of covariates, providing thus a promising approach to multivariate quantile regression.

Directional quantiles were proved to be continuous with respect to distributions and directions. We derived a closed-form expression of directional quantile envelopes for elliptically symmetric distributions. Directional quantile envelopes are affine equivariant and are capable of providing explicit probabilistic interpretation. We discussed the connection between directional quantile envelopes and halfspace depth contours and verified that distributions with smooth directional quantile envelopes are uniquely determined by the envelopes. We explored that under some conditions, the estimates of the directional quantile envelopes were consistent in the Pompeiu-Hausdorff distance sense. We also described the accuracy of the approximation of the estimates of the directional quantile envelopes.

The performance of our directional quantile regression approach was demonstrated by an application to the joint height-weight screening, bivariate growth charts. We

constructed bivariate growth charts and used them to locate individual subject's percentile rank. We also compared our directional approach to bivariate growth charts with the method proposed by Wei (2008). The two methods provided meaningful interpretation of the resulting envelopes (contours), which were complementary in scope: one by tangent mass and the other by enclosed mass. In the simulation study, both methods were capable of yielding envelopes (contours) which were a good approximation of the true ones. In the real data study, we demonstrated that our directional approach was flexible enough to add in covariates besides age, the way to obtain more informative readings than conventional univariate growth charts, BMI charts and Rohrer Index charts, as developed by Wei, Pere, Koenker and He (2005), Wei and He (2006), Wei (2008) (see also the discussion of Wei and He (2006)).

Besides developing the theoretical and practical aspects of multivariate quantile regression, we discussed some further problems in the theory of halfspace depth. A part of this was computation of depth contours, where we developed a fast elimination algorithm, to construct depth contours by eliminating nonactive directional quantile regression lines step by step. Applying this algorithm to a large number of quantile halfspaces, we can construct an arbitrary exact approximation of the direction quantile envelope.

We established the connection between halfspace depth contours and directional regression quantiles (Laine, 2001), stated without proof in Koenker (2005). Our proof uses the duality theory of primal-dual linear programming. We studied some properties of the structure of halfspace depth contours for empirical distributions, absolute continuous distributions and certain general distributions, which provided insight into explicit interpretation of depth contours.

We proposed a concept of generalized quantile, depth quantile, inspired by halfspace

depth (Tukey, 1975) and regression depth (Rousseeuw and Hubert, 1999). We explored its properties in various data-analytic cases. In the case of multivariate location, the depth quantiles are affine equivariance and half closed form was formalized when the distribution is elliptically symmetric. We found that they could be the usual quantiles in the special case of univariate location. In regression with intercept, it is identical to the original Koenker's depth-based regression quantile. We pointed out certain differences arising in regressions without intercept and gave an example showing that while the L_1-based quantile regression of Koenker and Bassett (1978) fails, our quantile regression provides a sensible answer.

Bibliography

[1] Abdous, B. and Theodorescu, R. (1992). Note on the spatial quantile of a random vector. *Statistics and Probability Letters*, 13, 333-336.

[2] Adrover, J., Maronna, R. A. and Yohai, V. J. (2004). Robust regression quantiles. *Journal of Statistical Planning Inference*, 122, 187-202.

[3] Appa, G. and Smith, C. (1973). On L_1 and Chebyshev estimation. *Mathematical Programming*, 5, 73-87.

[4] Aronson, J. K. (2001). Francis Galton and the invention of terms for quantiles. *Journal of Clinical Epidemiology*, 54, 1191-1194.

[5] Beirlant, J., Goegebeur, Y. Teugels, J. and Segers, J. (2004). *Statistics of Extremes: Theory and Applications.* Wiley, Chichester (with contributions from D. De Waal and C. Ferro).

[6] Bloomfield, P. and Steiger, W. L. (1983). *Least Absolute Deviations: Theory, Applications, and Algorithms*. Birkhäuser.

[7] Bosch, R. J., Y. Ye, and G. G. Woodworth (1995). A convergent algorithm for quantile regression with smoothing splines. *Computational Statistics & Data Analysis,* 19, 613-630.

[8] Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75, 761-771.

[9] Breckling, J., Kokic, P. and Lubke, O. (2001). A note on multivariate M-quantiles. *Statistics and Probability Letters*, 55, 39-44.

[10] Buchinsky, M. (1998). Recent advances in quantile regression models: a practical guideline for empirical research. *The Journal of Human Resources*, 33, 22-126.

[11] Cai, Z. and Xu, X. (2009). Nonparametric quantile estimations for dynamic smooth coefficient models. *Journal of the American Statistical Association*, 104, 371-383.

[12] Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics*, 53, 380-403.

[13] Chakraborty, B. (2003). On multivariate quantile regression. *Journal of Statistical Planning and Inference*, 110, 109-132.

[14] Chakraborty, B. and Chaudhuri, P. (1998). On an adaptive transformation-retransformation estimate of multivariate location. *Journal of the Royal Statistical Society, B,* 60, 145-157.

[15] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91, 862-872.

[16] Chemielewski, M. A. (1981). Elliptically symmetric distributions: a review and bibliography. *International Statistical Review*, 49, 67-74.

[17] Chen, C. (2005). An introduction to quantile regression and the QUANTREG procedure. *Proceedings of the Thirtieth Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.

[18] Cole, T. J. (1988). Fitting smoothed centile curves to reference data. *Journal of the Royal Statistics Society, A*, 151, 385-418.

[19] Cole, T. J. (1994). Growth charts for both cross-sectional and longitudinal data. *Statistics in Medicine*, 13, 2477-2492.

[20] Cole, T. J. and Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statistics in Medicine*, 11, 1305-1319.

[21] Cole, T. J., Freeman, J. V., and Preece, M. A. (1998). British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Statistics in Medicine*, 17, 407-429.

[22] Coles, S. G. and Tawn, J.A. (1991). Modelling extreme multivarite events. *Journal of the Royal Statistical Society, B,* 53, 377-392.

[23] Csorgo, M. (1983). Quantile Process with Statistical Applications. *Carleton University,* Society for Industrial and Applied Mathenatics.

[24] Deke, J. (2000). Simulataneous estimation of equations with quantile regression using Oja multivariate quantiles. *Unpublished (M.Sc.)Thesis*, University of Illinois, Urbana-Champaign.

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

[26] Edgeworth, F. Y. (1886). Problems in probabilities. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 5th series* 22, 371-384.

[27] Edgeworth, F. Y. (1893). Exercises in the calculation of errors. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 5th series* 36, 98-111.

[28] Efron, B. and Tibshirani, R. J. (1993). *An Introduction to Bootstrap*. Chapman and Hall, New York.

[29] Einmahl, J. H.J., and Mason, D.M. (1992). Generalized quantile processes. *Annals of Statistics,* 20, 1062-1078.

[30] Eubank, R. L. (1986). Quantiles. In *Encyclopedia of Statistical Sciences, Volume 7* (S. Kotz, N. L. Johnson and C. B. Read, eds). 424-432. Wiley, New York.

[31] Evans, M. (1982). Confidence bands for bivariate quantiles. *Communications in Statistics - Theory and Methods*, 11, 1465-1474.

[32] Fatti, L. P., Senaoana, E. M. and Thompson, M. L. (1998). Bayesian updating in reference centile charts. *Journal of Royal Statistical Society, A*, 161, 103-115.

[33] Ferguson, T. S. (1967). *Mathematical Statistics: A Desicison Theoretic Approach.* Academic Press, New York.

[34] Frohne, I. and Hyndman, R. J. (2004). `quantile`. A function in R starting from version 2.0.0, `http://www.r-project.org`.

[35] Galton, F. (1888-1889). Corelations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45, 135-145.

[36] Gilchrist, W. (2000). *Statistical Modelling with Quantile Functions.* Chapman & Hall/CRC, New York.

[37] Goldberger, A.S. (1968). *Topics in Regression Analysis*. Macmillan, New York.

[38] Hajek, J. and Vorlickova, D. (1977). *Mathematical Statistics*. SPN, Praha [in Czech].

[39] Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika*, 35, 414-415.

[40] Hamill, P.V.V., Dridzd, T.A., Johnson, C.L., Reed, R.B., Roche,A.F. and Moore, W.M. (1979). Physical growth: National Center for Health Statistics percentiles. *American Journal of Clinical Nutrition*, 32, 607-629.

[41] Hassairi, A. and Regaieh, O. (2007). On the Tukey depth of an atomic measure. *Statistical Methodology*, 4, 244-249.

[42] He, X. and Wang, G. (1997). Convergence of depth contours for multivariate datasets. *The Annals of Statistics*, 25, 495-504.

[43] Hodges, J. L. Jr. (1955). A bivariate sign test. *The Annals of Mathematical Statistics,* 26, 523-527.

[44] Hogg, V. R. (1975). Estimates of percentile regression lines using salary data. *Journal of the American Statistical Association*, 70, 56-59.

[45] Hosking, J. R. M., Wallis, J. R., and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27, 251-261.

[46] Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics,* 35, 73-101.

[47] Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, 50, 361-365.

[48] Jackel, P. (2002). *Monte Carlo Methods in Finance.* Wiley, New York.

[49] Kim, M. O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35, 92-108.

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

[51] Koenker, R. (2007). `Quantreg`: an R package for quantile regression and relative methods. `http://cran.r-project.org`.

[52] Koenker, R. and Bassett, G. W., Jr. (1978). Regression quantiles. *Econometrica*, 46, 33-50.

[53] Koenker, R. and Hallock, K. (2001). Quantile regression. *The Journal of Economic Perspectives*, 15, 143-156.

[54] Koenker, R. and Mizera, I. (2004). Penalized triograms: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society, B*, 66, 145-163.

[55] Koenker, R., Ng, P. and Portnoy, S. (1994), Quantile smoothing splines. *Biometrika,* 81, 673-680.

[56] Koltchinskii, V. (1994). Spatial quantiles and their Bahadur-Kiefer representations. In *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium,* held from September 4-9, 1993. (P. Mandl and M. Huskova, eds.) 361-367. Physica-Verlag, Heidelberg.

[57] Koltchinskii, V. (1996). M-estimation and spatial quantiles. In *Robust statistics, data analysis, and computer intensive methods: in honor of Peter Huber's 60th birthday* (H. Rieder, ed.) 235-250. Springer, New York.

[58] Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25, 435-477.

[59] Koshevoy, G. A. (2001). Projections of lift zonoids, the Oja depth and Tukey depth. *Preprint*.

[60] Koshevoy, G. A. (2002). The Tukey depth characterizes the atomic measure. *Journal of Multivariate Analysis*, 83, 360-364.

[61] Laine, B. (2001). Depth contours as multivariate quantiles: a directional approach. *Master thesis.* Universite Libre de Bruxelles.

[62] Ledford, A. W. and Tawn, J.A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83, 169-187.

[63] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *The Annals of Statistics*, 23, 783-840.

[64] Lucchetti, R., Torre, A. and Wets, R. J.-B. (1993). Uniform convergence of probability measures: topological criteria. *Canadian Mathematical Bulletin*, 36, 197-208.

[65] Lucchetti, R., Salinetti, G. and Wets, R. J.-B. (1994). Uniform convergence of probability measures: topological criteria. *Journal of Multivariate Analysis*, 51, 252-264.

[66] Manski, C. F. (1988). *Analog Estimation Methods in Econometrics*. Chapman and Hall, New York.

[67] Mardia, L. V., Kent, J. T. and Bibby, J. M. (1979), *Multivariate Analysis.* Academic Press, New York.

[68] Massé, J. -C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10, 397-419.

[69] Massé, J. -C. and Theodorescu, R. (1994). Halfplane trimming for bivariate distributions. *Journal of Multivariate Analysis*, 48, 188-202.

[70] McNeil, A. and Stephens, A. (2007). `evir`: *Extreme Values in R*. S original (EVIS) by Alexander McNeil and R port by Alec Stephens, R package version 1.5, `http://www.math.lancs.ac.uk/~stephena/`.

[71] Mizera, I. (2002). On depth and deep points: A calculus. *The Annals of Statistics*, 30, 1681-1736.

[72] Mizera, I. and Müller, C. H. (2004). Location-scale depth (with discussion). *Journal of the American Statistical Association*, 99, 949-966.

[73] Mizera, I. and Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *Journal of Multivariate Analysis*, 83, 365-388.

[74] Nadar, M., Hettmansperger, T.P., and Oja H., (2003). The asymptotic variance of the Oja median. *Statistics and Probability Letters*, 64, 431-442.

[75] Nash, S. G. and Sofer, A (1996). *Linear and Nonlinear Programming*. McGraw-Hill: New York.

[76] Ng, P. T. and Maechler, M. (2006). `cobs`: COBS-Constrained B-splines (Sparse matrix based). R package version 1.1-3.5, `http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:cobs`.

[77] Nolan, D. (1992). Asymptotics for multivariate trimming. *Stochastic Processes and their Applications*, 42, 157-169.

[78] Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1, 327-332.

[79] Parzen, E. (2004). Quantile probability and statistical data modelling. *Statistical Science,* 19, 652-662.

[80] Pere, A. (2000). Comparison of two methods of transforming height and weight to normality. *The Annals of Human Biology,* 27, 35-45.

[81] Pere, A. (2006). Comments on Conditional growth charts by Y. Wei and X. He. *The Annals of Statistics*. 34, 2105-2112.

[82] Petersen, J. H. (2003), Two bivariate geometrically defined reference region with application to male reproductive hormones and human growth. *Statistics in Medicine*, 22, 2603-2618.

[83] Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3, 119-131.

[84] R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`

[85] Reiss, R. D. and Thomas, M. (2007). *Statistical Analysis of Extreme Values.* Birkhauser Verlag, Basel.

[86] Resnick, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer, New York.

[87] Rockafellar, R. T. (1996). *Convex Analysis*. Princeton University Press, Princeton.

[88] Rockafellar, R. T. and Wets, R. J. -B. (1998). *Variational Analysis.* Springer-Verlag, Berlin.

[89] Ronkainen, T., Oja, H. and Orponen, P. (2001). Computation of the multivariate Oja median. *Proceedings of International Conference on Robust Statistics*, 344-359.

[90] Rousseeuw, P. J. and Hubert, M. (1999). Regression depth (with discussion). *Journal of the American Statistical Association*, 94, 388-402.

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

[92] Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49, 213-244.

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

[94] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003).*Semiparametric Regression.* Cambridge University Press, Cambridge.

[95] Schneider, R. (1993). *Convex bodies: The Brunn-Minkowski theory*. Cambridge University Press, Cambridge.

[96] Serfling, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica*, 56, 214-232.

[97] Shorack, G. R. (2000). *Probability for Statisticians*, Springer-Verlag, New York.

[98] Small, G. C. (1990). A survey of multidimensional medians. *International Statistical Review*, 58, 263-277.

[99] Sposito, V. A. (1975). *Linear and Nonlinear Programming*. Ames, Iowa State University Press.

[100] Stephenson, A. and Gilleland, E. (2006). Software for the analysis of extreme events: the current state and future directions. *Extremes*, 8, 87-109.

[101] Struyf, A. and Rousseeuw, P. J. (1999). Halfspace depth and regression depth characterize the empirical distribution. *Journal of Multivariate Analysis*, 69, 135-153.

[102] Thompson, M.L., and Fatti, L.P. (1997). Construction of multivariate centile charts for longitudianl measurements. *Statistics in Medicaine*, 16, 333-345.

[103] Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 2, 523-531.

[104] Walther, G. (1997a). Monte Carlo sampling in dual space for approximating the empirical halfspace distance. *The Annals of Statistics*, 25, 1926-1953.

[105] Walther, G. (1997b). On a conjecture concerning a theorem of Cramér. *Journal of Multivariate Analysis*, 63, 313-319.

[106] Whittaker, E. (1923). On a new method of graduation. *Proceedings Of Edingburgh Mathematical Society,* 41, 63-75.

[107] Wei, Y. (2008). An approach to multivariate covariate-dependent quantile contours with application to bivariate conditional growth charts. *Journal of the American Statistical Association*, 103, 397-409.

[108] Wei, Y. and He, X. (2006). Conditional growth charts (with discussion). *The Annals of Statistics,* 34, 2069-2097.

[109] Wei, Y., Pere, A., Koenker, R. and He, X. (2006). Quantile regression methods for reference growth charts. *Statistics in Medicine*, 25, 1369-1382.

[110] Wolf, P. H. (2007). Another Plot PACKage: stem.leaf, bagplot, faces, spin3R, ... `http://cran.r-project.org`.