

REDUCING THE SAMPLE SIZE WHEN ESTIMATING CONDITIONAL QUANTILES

S. Shatskikh¹, L. Melkumova²

¹Samara National Research University, Samara, Russia

²Mercury Development Russia, Samara, Russia

Abstract. In this paper we propose an idea to use a certain property of multivariate probability distributions, that we call the conditional quantile reproducibility, to decrease the amount of observations required to construct a statistical estimate of a n-dimensional conditional quantile of the distribution. For the class of probability distributions, satisfying to this property, we present several results, proving that in many cases the reproducibility property allows us to restore the n-dimensional conditional quantile by solving a certain type of Pfaffian differential equation. The equation is constructed from functions, derived only from the 2-dimensional marginal distributions of the initial distribution.

Keywords: multivariate conditional quantiles, quantile reproducibility, Pfaffian quantile differential equation, conditional quantile estimation

Citation: Shatskikh S, Melkumova L. Reducing the Sample Size When Estimating Conditional Quantiles. CEUR Workshop Proceedings, 2016; 1638: 769-781. DOI: 10.18287/1613-0073-2016-1638-769-781

1 Introduction

Many of the widely used multivariate probability distributions satisfy to a certain property of the multivariate conditional quantiles which we call “quantile reproducibility”. Among them are the multivariate Gaussian distribution, Student’s distribution, Logistic distribution, Pareto distribution, Gamma distribution, Clayton’s Copula distribution and others (see [9, 8, 6, 5]).

This article gives several results related to the quantile reproducibility property of the multivariate distributions. Particularly we prove that when the distribution satisfies to the quantile reproducibility property (we also say it “has reproducible conditional quantiles”), the solution of a Pfaffian differential equation of certain form, that can be constructed by only knowing 2-dimensional marginal distributions of the original distribution, is equal to a distribution’s multivariate conditional quantile. In other words by solving the Pfaffian differential equation we build a multivariate quantile from bivariate functions, describing the probability distribution.

The outline of this article is as follows. In Section 2 we introduce the multivariate conditional quantiles and discuss some of their applications. In Section 3 we give

the definition of the quantile reproducibility for multivariate distributions and talk on its geometrical interpretation.

In Section 4 we introduce the Pfaffian differential equation for the distribution which we refer to as the “quantile equation”. We also show that the solution of the maximum possible dimension for the quantile equation of a distribution with reproducible quantiles equals a conditional quantile of the maximum dimension. Section 5 gives several concrete examples of multivariate distributions with reproducible quantiles along with the corresponding quantile equations.

In Section 6 we give an intermediate version of the quantile reproducibility property and prove that, when it is satisfied, one can find the solutions of the quantile equation of intermediate dimensions. In Section 7 we illustrate this theorem by giving an example of the distribution with this version of quantile reproducibility and finding the solutions of the quantile equation for it.

In Section 8 we discuss how the quantile reproducibility property could be used in statistical estimation. For the case when it is known that the distribution has reproducible conditional quantiles, we propose a technique to build the multivariate quantile estimate by only using bivariate observations. We also show that this technique allows us to reduce the required number of observations when compared to a traditional approach to multivariate quantile estimation.

2 Conditional Quantiles and Their Applications

First, we will give the definition of the multivariate conditional quantile. Consider a system of random variables X_1, X_2, \dots, X_n .

Suppose we have the conditional cumulative distribution function

$$F_{i|1\hat{\dots}i\dots n}(x_i|x_1, \dots, \hat{x}_i, \dots, x_n)$$

(the sign $\hat{\cdot}$ over the element \cdot means that the element is omitted), which is continuous and increases monotonically in x_i for any fixed vector $(x_1, \dots, \hat{x}_i, \dots, x_n) \in \mathbb{R}^{n-1}$.

The *conditional quantile* $q_{i|1\hat{\dots}i\dots n}^{(p)}(x_1, \dots, \hat{x}_i, \dots, x_n)$ of level $p \in [0, 1]$ for a random variable X_i given $X_1, \dots, \hat{X}_i, \dots, X_n$ is defined by the equation

$$F_{i|1\hat{\dots}i\dots n}(q_{i|1\hat{\dots}i\dots n}^{(p)}(x_1, \dots, \hat{x}_i, \dots, x_n)|x_1, \dots, \hat{x}_i, \dots, x_n) \equiv p$$

for any $(x_1, \dots, \hat{x}_i, \dots, x_n) \in \mathbb{R}^{n-1}$.

The *conditional median* $m_{i|1\hat{\dots}i\dots n}(x_1, \dots, \hat{x}_i, \dots, x_n)$ for a random variable X_i given $X_1, \dots, \hat{X}_i, \dots, X_n$ is a conditional quantile of level $p = \frac{1}{2}$.

Another way to define a conditional quantile is to give a certain point $\mathbf{x}^\circ = (x_1^\circ, \dots, x_n^\circ) \in \mathbb{R}^n$, lying on the quantile surface:

$$F_{1|2\dots n}\left(q_{1|2\dots n}^{(\mathbf{x}^\circ)}(x_2, \dots, x_n)|x_2, \dots, x_n\right) \equiv F_{1|2\dots n}(x_1^\circ|x_2^\circ, \dots, x_n^\circ).$$

The conditional quantile is then said to be *going through* \mathbf{x}° . In this case the level of the quantile is given by $p = F_{1|2\dots n}(x_1^\circ|x_2^\circ, \dots, x_n^\circ)$.

One of the several applications of the conditional quantiles is using them as random variable estimates. Consider the following situation. We have a set of random variables X_1, X_2, \dots, X_n . We observe the first $n - 1$ of them, and

the last one needs to be estimated. We also know the conditional distribution function $F_{n|1\dots n-1}(x_n | x_1, \dots, x_{n-1})$ which is supposed to be monotonic in x_n . Now, having defined a loss function

$$\rho_\tau(u) = \begin{cases} (\tau - 1)u, & u \leq 0, \\ \tau u, & u > 0. \end{cases} \quad \tau \in (0, 1).$$

we want to get the estimate for the last random variable X_n , so that it would minimize the conditional loss.

$$\hat{f}(x_1, \dots, x_{n-1}) : \min_{f(\cdot)} \mathbb{M}\{\rho_\tau(X_n - f(x_1, \dots, x_{n-1})) | x_1, \dots, x_{n-1}\}$$

Rather simple calculations will lead us to the following formula for the X_n estimate.

$$F_{n|1\dots n-1}(\hat{f}(x_1, \dots, x_{n-1}) | x_1, \dots, x_{n-1}) \equiv \tau$$

That is any value to satisfy to this equation would minimize the loss function. As long as the F is monotonic in x_n , there is only one such value, which is the conditional quantile

$$\hat{f}(x_1, \dots, x_{n-1}) = F_{n|1\dots n-1}^{-1}(\tau | x_1, \dots, x_{n-1}) = q_{n|1\dots n-1}^{(\tau)}(x_1, \dots, x_{n-1}).$$

Further on, it is also known (see [10]), that the estimate, minimizing the conditional loss, will also minimize the expected loss (the bayesian risk):

$$\mathbb{M}\{\rho_\tau(X_n - \hat{f}(X_1, \dots, X_{n-1}))\} = \min_{f(\cdot)} \mathbb{M}\{\rho_\tau(X_n - f(X_1, \dots, X_{n-1}))\}$$

3 Reproducible Conditional Quantiles

Consider a vector of random variables $\mathbf{X} = (X_1, \dots, X_n)$ with the cumulative distribution function $F_{1\dots n}(x_1, \dots, x_n)$, the multivariate strictly positive density

$$f_{1\dots n}(x_1, \dots, x_n) > 0, \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

and the conditional distribution functions

$$\mathbb{P}\{X_1 \leq x_1 | X_2 = x_2, \dots, X_n = x_n\} = F_{1|2\dots n}(x_1 | x_2, \dots, x_n),$$

$$\mathbb{P}\{X_i \leq x_i | X_j = x_j\} = F_{i|j}(x_i | x_j), \quad i \neq j, \quad i, j = \overline{1, n}.$$

By fixing a point $\mathbf{x}^\circ = (x_1^\circ, \dots, x_n^\circ) \in \mathbb{R}^n$, we get a family of conditional quantiles, going through the selected point, which act as level surfaces (or curves) of the conditional CDF:

$$F_{1|2\dots n}(q_{1|2\dots n}^{(\mathbf{x}^\circ)}(x_2, \dots, x_n) | x_2, \dots, x_n) \equiv F_{1|2\dots n}(x_1^\circ | x_2^\circ, \dots, x_n^\circ),$$

$$q_{1|2\dots n}^{(\mathbf{x}^\circ)}(x_2^\circ, \dots, x_n^\circ) = x_1^\circ,$$

$$F_{i|j}(q_{i|j}^{(x_i^\circ, x_j^\circ)}(x_j) | x_j) \equiv F_{i|j}(x_i^\circ | x_j^\circ), \quad q_{i|j}^{(x_i^\circ, x_j^\circ)}(x_j^\circ) = x_i^\circ, \quad i \neq j.$$

To simplify the notation let us expand the determinant along the first row

$$\mathbf{W}(\mathbf{x}^\circ) = \sum_{k=1}^n A_{1k}(x_1^\circ, \dots, x_n^\circ) \mathbf{e}_k.$$

Each of the cofactors A_{1k} will depend on a set of quantile derivatives $\dot{q}_{i|j}^{(x_i^\circ, x_j^\circ)}(x_j^\circ)$. Now we replace the point $\mathbf{x}^\circ = (x_1^\circ, \dots, x_n^\circ)$ with $\mathbf{x} = (x_1, \dots, x_n)$, which is variable in \mathbb{R}^n , and consider the differential 1-form

$$\omega = \sum_{k=1}^n A_{1k}(x_1, \dots, x_n) dx_k.$$

Next we construct a Pfaffian differential equation for the form

$$\omega = \sum_{k=1}^n A_{1k}(x_1, \dots, x_n) dx_k = 0.$$

We call it the *quantile equation*.

Theorem 1. *If the probability distribution $F_{1\dots n}(x_1, \dots, x_n)$ with a joint PDF positive on \mathbb{R}^n has reproducible conditional quantiles (1), and $A_{11}(x_1, \dots, x_n) \neq 0$, then the quantile equation*

$$\omega = \sum_{i=1}^n A_{1i}(x_1, \dots, x_n) dx_i = 0 \quad (2)$$

is completely integrable. The solution of (2) going through the given point \mathbf{x}° is the “big” conditional quantile $x_1 = q_{1|2\dots n}^{(\mathbf{x}^\circ)}(x_2, \dots, x_n)$.

The proof of this theorem is given in [6].

A well known result, the Frobenius theorem (see [1] p. 97), gives a necessary and sufficient condition of the complete integrability of the Pfaffian differential equation. It states that the equation (2) is completely integrable if and only if

$$d\omega \wedge \omega = 0, \quad (3)$$

where $d\omega$ is the exterior differential of the differential 1-form ω and \wedge means the exterior product of the two differential forms.

5 Examples

Many commonly used multivariate distributions have reproducible conditional quantiles. Some of the examples are multivariate Gaussian distribution, multivariate Gamma distribution, multivariate Student distribution, multivariate Logistic distribution, multivariate Pareto distribution, Clayton copula (see [4]). Here to illustrate our results we present only a few specific distributions along with their quantile equations. All of the equations can be solved using the

well-known elementary methods (see for example [7]).

Tabl.1. Quantile equations for some distributions

Densities	Quantile equations
1. $\phi(\mathbf{x}; \mathbf{m}, [\sigma^{ij}])$	$\sum_{k=1}^n \sigma^{1k} dx_k = 0$
2. $\frac{1}{\pi^2(1+x_1^2+x_2^2+x_3^2)^2}$	$(1+x_2^2+x_3^2)dx_1 - x_1x_2dx_2 - x_1x_3dx_3 = 0$
3. $\frac{6e^{-(x_1+x_2+x_3)}}{(1+e^{-x_1}+e^{-x_2}+e^{-x_3})^4}$	$(e^{x_2}+e^{x_3}+e^{x_2+x_3})dx_1 - e^{x_3}dx_2 - e^{x_2}dx_3 = 0$
4. $\frac{6}{(x_1+x_2+x_3-2)^4}$	$(x_2+x_3-1)dx_1 + (1-x_1)dx_2 + (1-x_1)dx_3 = 0$
$x_i \geq 1, i = 1, 2, 3$	

1–Gaussian distribution, 2–Cauchy distribution, 3–Logistic distribution,
4–Pareto distribution

6 The Darboux Class for Quantile Differential Equations and Its Relation to Reproducibility

Now what if the quantile differential equation is not completely integrable? Are there cases when we can say something about the solutions of this equation?

To answer this question, let us first remind the reader about one of the characteristics of the differential 1-forms (and as a consequence of the Pfaffian differential equations) - the Darboux class. As the Darboux theorem (see [3]) states, the class gives the maximum dimension of the integral manifold of the corresponding Pfaffian differential equation.

Definition 2. *If the differential 1-form ω satisfies the equality*

$$\omega \wedge (d\omega)^r \neq 0, \quad \text{but} \quad \omega \wedge (d\omega)^{r+1} = 0,$$

then we say that the Darboux class of the differential form equals $2r + 1$.

As we have already mentioned (see section 4), the equality

$$d\omega \wedge \omega = 0 \tag{4}$$

gives the criterion for complete integrability of a Pfaffian equation (Frobenius theorem). In this case $r = 0$, so the Darboux class of the differential form ω is equal to 1.

Theorem (Darboux's theorem). *If the Darboux class of the differential 1-form ω equals $2r + 1$, then by a smooth local change of coordinates the Pfaffian equation*

$$\omega = \sum_{k=1}^n a_k(x_1, \dots, x_n) dx_k = 0$$

¹with coefficients $a_k(x_1, \dots, x_n)$ not equal to zero simultaneously

can be converted to the canonical form

$$dy_1 + y_2 dy_3 + \dots + y_{2r} dy_{2r+1} = 0.$$

In this case the Pfaffian equation has an integral manifold of maximum dimension $n - r - 1$ with the following first integrals:

$$y_1(x_1, \dots, x_n) = C_1 = \text{const}, \quad y_3(x_1, \dots, x_n) = C_3 = \text{const}, \dots, \\ y_{2r+1}(x_1, \dots, x_n) = C_{2r+1} = \text{const}.$$

As noted above, if the condition (4) holds, then $r = 0$. So, according to the Darboux's theorem, the maximum dimension of the integral manifold must equal $n - 1$, which means the complete integrability of the equation. This agrees with the Frobenius theorem given earlier.

Let us now sum up. For the given quantile equation we can calculate the Darboux class of the corresponding differential 1-form. From this value we obtain the maximum dimension of the integral manifold. But still what is the integral manifold itself?

We will now show, that for multivariate probability distributions with certain type of quantile reproducibility, the solutions can be given explicitly.

To do this, we will first establish one useful property of matrix determinants (see [5]).

Let us consider the determinant of order n , where $n - 1 \geq k$, of the form²

$$M = \begin{vmatrix} dx_1 & dx_2 & \dots & dx_k & dx_{k+1} & \dots & dx_{n-1} & dx_n \\ 1 & \dot{q}_{2|1} & \dots & \dot{q}_{k|1} & \dot{q}_{k+1|1} & \dots & \dot{q}_{n-1|1} & \dot{q}_{n|1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dot{q}_{1|k} & \dot{q}_{2|k} & \dots & 1 & \dot{q}_{k+1|k} & \dots & \dot{q}_{n-1|k} & \dot{q}_{n|k} \\ \dot{q}_{1|k+1} & \dot{q}_{2|k+1} & \dots & \dot{q}_{k|k+1} & 1 & \dots & \dot{q}_{n-1|k+1} & \dot{q}_{n|k+1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dot{q}_{1|n-1} & \dot{q}_{2|n-1} & \dots & \dot{q}_{k|n-1} & \dot{q}_{k+1|n-1} & \dots & 1 & \dot{q}_{n|n-1} \end{vmatrix}.$$

We will denote the cofactor of the element dx_i in the first row of M by $A(dx_i)$.

Lemma 1. *The following expansion for the determinant M is true:*

$$\frac{A(dx_{k+1})}{S} \begin{vmatrix} dx_1 & dx_2 & \dots & dx_k & dx_{k+1} \\ 1 & \dot{q}_{2|1} & \dots & \dot{q}_{k|1} & \dot{q}_{k+1|1} \\ \dots & \dots & \dots & \dots & \dots \\ \dot{q}_{1|k} & \dot{q}_{2|k} & \dots & 1 & \dot{q}_{k+1|k} \end{vmatrix} + \dots + \frac{A(dx_n)}{S} \begin{vmatrix} dx_1 & dx_2 & \dots & dx_k & dx_n \\ 1 & \dot{q}_{2|1} & \dots & \dot{q}_{k|1} & \dot{q}_{n|1} \\ \dots & \dots & \dots & \dots & \dots \\ \dot{q}_{1|k} & \dot{q}_{2|k} & \dots & 1 & \dot{q}_{n|k} \end{vmatrix}, \quad (5)$$

where

$$S = \begin{vmatrix} 1 & \dot{q}_{2|1} & \dots & \dot{q}_{k|1} \\ \dots & \dots & \dots & \dots \\ \dot{q}_{1|k} & \dot{q}_{2|k} & \dots & 1 \end{vmatrix}.$$

The proof of this lemma is given in [5].

²The given expansion holds for determinants of general form.

that is

$$w_i(x_1, \dots, x_k, q_{i|1\dots k}^{(x^0)}(x_1, \dots, x_k)) \equiv 0, \quad i = \overline{k+1, n}. \tag{10}$$

Let us now consider the quantile equation for the initial distribution.

$$w(x_1, \dots, x_n) = \begin{vmatrix} dx_1 & \dots & dx_k & \dots & dx_{n-1} & dx_n \\ 1 & \dots & \dot{q}_{k|1}^{(x^0)}(x_1) & \dots & \dot{q}_{n-1|1}^{(x^0)}(x_1) & \dot{q}_{n|1}^{(x^0)}(x_1) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dot{q}_{1|n-1}^{(x^0)}(x_{n-1}) & \dots & \dot{q}_{i|n-1}^{(x^0)}(x_{n-1}) & \dots & 1 & \dot{q}_{n|n-1}^{(x^0)}(x_{n-1}) \end{vmatrix} = 0.$$

We can apply Lemma 1 to expand the left part of the equation:

$$w(x_1, \dots, x_n) = \frac{A(dx_{k+1})}{S} \cdot w_{k+1}(x_1, \dots, x_k, x_{k+1}) + \dots + \frac{A(dx_n)}{S} \cdot w_n(x_1, \dots, x_k, x_n).$$

Now, using (10), we get:

$$w(x_1, \dots, x_k, q_{k+1|1\dots k}^{(x^0)}(x_1, \dots, x_k), \dots, q_{n|1\dots k}^{(x^0)}(x_1, \dots, x_k)) \equiv 0.$$

Therefore, the surface (8) is an integral manifold for the initial quantile equation (2). □

Note 1. If the conditions of the theorem are satisfied, the quantile equation (2) has a solution of dimension k . Therefore the maximum possible integral manifold dimension for the equation is not less than k . And, consequently, the Darboux class of the 1-form ω is less or equal to $2(n - k) - 1$.

When the Darboux class of the quantile equation is equal to $2(n - k) - 1$, the surface (8) is the integral manifold of (2) of maximum possible dimension, going through the point x^0 .

Note 2. If we add to (6) the following condition

$$q_{n|1\dots n-1}^{(x^0)}(x_1, \dots, x_k, q_{k+1|1\dots k}^{(x^0)}(x_1, \dots, x_k), \dots, q_{n-1|1\dots k}^{(x^0)}(x_1, \dots, x_k)) \equiv q_{n|1\dots k}^{(x^0)}(x_1, \dots, x_k),$$

then the integral manifold (8) takes the form:

$$\left\{ \left(x_1, \dots, x_k, q_{k+1|1\dots k}^{(x^0)}(x_1, \dots, x_k), \dots, q_{n-1|1\dots k}^{(x^0)}(x_1, \dots, x_k), q_{n|1\dots n-1}^{(x^0)} \left(x_1, \dots, x_k, q_{k+1|1\dots k}^{(x^0)}(x_1, \dots, x_k), \dots, q_{n-1|1\dots k}^{(x^0)}(x_1, \dots, x_k) \right) \right) \right\}$$

and it is a part of the “big” conditional quantile surface

$$\left\{ \left(x_1, x_2, \dots, x_{n-1}, q_{n|1\dots n-1}^{(x^0)}(x_1, \dots, x_{n-1}) \right) \right\}.$$

7 Example of the Distribution with an Intermediate Darboux Class

To illustrate theorem 2, let us consider a mixture of two 5-dimensional Cauchy distributions with density

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{1}{3}c^{(1)}(x_1, x_2, x_3, x_4, x_5) + \frac{2}{3}c^{(2)}(x_1, x_2, x_3, x_4, x_5),$$

where

$$c^{(1)} = \frac{12\sqrt{15}}{\pi^3 (1 + x_1^2 + 3x_2^2 + 4x_3^2 + \frac{45}{2}x_4^2 - 15x_4x_5 + \frac{9}{2}x_5^2)^3},$$

$$c^{(2)} = \frac{40\sqrt{3}}{\pi^3 (1 + x_1^2 + 3x_2^2 + 4x_3^2 + 50x_4^2 + 40x_4x_5 + 10x_5^2)^3}.$$

We have the following quantile equation

$$\omega = \begin{vmatrix} dx_1 & dx_2 & dx_3 & dx_4 & dx_5 \\ 1 & \frac{x_1x_2}{1+(x_1)^2} & \frac{x_1x_3}{1+(x_1)^2} & \frac{x_1x_4}{1+(x_1)^2} & \frac{x_1x_5}{1+(x_1)^2} \\ \frac{3x_1x_2}{1+3(x_2)^2} & 1 & \frac{3x_2x_3}{1+3(x_2)^2} & \frac{3x_2x_4}{1+3(x_2)^2} & \frac{3x_2x_5}{1+3(x_2)^2} \\ \frac{4x_1x_3}{1+4(x_3)^2} & \frac{4x_2x_3}{1+4(x_3)^2} & 1 & \frac{4x_3x_4}{1+4(x_3)^2} & \frac{4x_3x_5}{1+4(x_3)^2} \\ \frac{10x_1x_4}{1+10(x_4)^2} & \frac{10x_2x_4}{1+10(x_4)^2} & \frac{10x_3x_4}{1+10(x_4)^2} & 1 & \dot{q}_{5|4}^{(x_4, x_5)}(x_4) \end{vmatrix} = 0, \tag{11}$$

where

$$\dot{q}_{5|4}^{(x_4, x_5)}(x_4) = A/B;$$

$$A = (1 + 10x_4^2)^{-1} \cdot \left(2\sqrt{10}(-1 + 5x_4x_5) (2 + 45x_4^2 - 30x_4x_5 + 9x_5^2)^{3/2} + 5(1 + 6x_4x_5) (1 + 10(5x_4^2 + 4x_4x_5 + x_5^2))^{3/2} \right);$$

$$B = \sqrt{10} (2 + 45x_4^2 - 30x_4x_5 + 9x_5^2)^{3/2} + 3 (1 + 10(5x_4^2 + 4x_4x_5 + x_5^2))^{3/2}.$$

Now let us calculate the Darboux class of the form ω to determine the maximum dimension of the solution of the quantile equation. The calculations show that

$$d\omega \neq 0; \quad d\omega \wedge \omega \neq 0 \text{ almost surely in } \mathbb{R}^5; \quad d\omega \wedge d\omega \equiv 0.$$

So the Darboux class of the form ω is almost surely equal to 3 and the maximum dimension of the solution of the quantile equation (11) is also almost surely equal to 3.

The integral manifold of the maximum possible dimension, going through the point $\mathbf{x}^\circ = (x_1^\circ, \dots, x_5^\circ)$ is given by the equalities

$$x_4 = x_4^\circ \sqrt{\frac{1 + x_1^2 + 3x_2^2 + 4x_3^2}{1 + (x_1^\circ)^2 + 3(x_2^\circ)^2 + 4(x_3^\circ)^2}}; \quad x_5 = x_5^\circ \sqrt{\frac{1 + x_1^2 + 3x_2^2 + 4x_3^2}{1 + (x_1^\circ)^2 + 3(x_2^\circ)^2 + 4(x_3^\circ)^2}},$$

which exactly match the two 3-dimensional conditional quantiles going through \mathbf{x}° : $\dot{q}_{4|123}^{(x^\circ)}(x_1, x_2)$ and $\dot{q}_{5|123}^{(x^\circ)}(x_1, x_2)$. That is the solution is

$$\mathcal{S}(\mathbf{x}^\circ) = \left\{ \left(x_1, x_2, x_3, \dot{q}_{4|123}^{(x^\circ)}(x_1, x_2, x_3), \dot{q}_{5|123}^{(x^\circ)}(x_1, x_2, x_3) \right) \right\}.$$

Finally it is easy to verify that the 3-dimensional conditional quantiles satisfy to the quantile reproducibility property. So, according to theorem 2, $\mathcal{S}(\mathbf{x}^\circ)$ should be the solution of the quantile equation. Since the Darboux class of the form ω is equal to 3, then, as it is stated in note 1, this is the solution of the maximum possible dimension.

It is also easy to show, that the condition of note 2 is satisfied, so $\mathcal{S}(\mathbf{x}^\circ)$ is a part of the “big” conditional quantile surface

$$\mathcal{S}(\mathbf{x}^\circ) \subseteq \left\{ \left(x_1, x_2, x_3, x_4, q_{5|1234}^{(\mathbf{x}^\circ)}(x_1, x_2, x_3, x_4) \right) \right\}.$$

8 Statistical Application of Quantile Reproducibility

The theorem 1 talks on how to obtain the $n - 1$ -dimensional quantile from a set of 1-dimensional quantile derivatives, which obviously can be calculated from bivariate marginal densities of the distribution. This logically leads to an idea to try to build a statistical estimate of the “big” conditional quantile of the distribution from a number of estimates of its bivariate densities. The outline of the algorithm would be as follows:

- First estimate 1-dimensional quantiles of the distribution and their derivatives from a number of bivariate observations (see for example [2]).
- Then using these quantile derivative estimates build the Pfaffian quantile equation 2 so that it's completely integrable and its solution approximates the “big” conditional quantile.
- Solve the quantile equation numerically and obtain the n -dimensional quantile estimate.

Obviously for this algorithm one only needs a set of 2-dimensional observations all of which can be made independently. With the direct approach to estimate the $n - 1$ -dimensional quantile one would need to observe the entire vector of n dimensions.

Now let us roughly compare the number of observations required to build the $n - 1$ -dimensional quantile estimate when using the traditional direct approach and the algorithm described above. For convenience we will assume that the quantile and the quantile derivative estimation in both cases is done by first estimating the corresponding distribution densities and deriving conditional densities from the estimates.

With the traditional approach we consider a sample of observations

$$\left\{ (x_1^{(1)}, \dots, x_n^{(1)}), \dots, (x_1^{(r_n)}, \dots, x_n^{(r_n)}) \right\}$$

for a random vector $\mathbf{X} = (X_1, \dots, X_n)$ with the density $f_{1\dots n}(x_1, \dots, x_n)$. From the observations we construct the density estimate $\hat{f}_{1\dots n}(x_1, \dots, x_n)$.

To simplify calculations we'll suppose that the histogram method is used. For each variable X_i we divide the sample range into m intervals. This way we get m^n n -dimensional parallelepipeds. If, in order to get a good estimate in each parallelepiped we need k observations, then the total amount of observations required to estimate the density of \mathbf{X} would be $r_n = k \cdot m^n$.

If we use the algorithm proposed above we don't need to estimate the n -variate density $f_{1\dots n}(x_1, \dots, x_n)$. Instead we construct estimates for the marginal densities $f_{ij}(x_i, x_j)$, $i \neq j$, $i, j = \overline{1, n}$, using the observations

$$\left\{ (x_i^{(1)}, x_j^{(1)}), \dots, (x_i^{(r_2)}, x_j^{(r_2)}) \right\}.$$

If again for a good estimate we need k observations in each of the rectangles, then each of the density estimates $\hat{f}_{ij}(x_i, x_j)$ will require $r_2 = km^2$ observations. And the total number of observations required to estimate all the bivariate densities equals

$$s_n = r_2 \cdot \frac{n \cdot (n-1)}{2} = km^2 \cdot \frac{n \cdot (n-1)}{2}.$$

It's clear that for $n \geq 3$

$$\lim_{m \rightarrow \infty} \frac{r_n}{s_n} = \infty,$$

which means that the overall number of observations required to construct a n -dimensional quantile when n is relatively big would be much less in case if the proposed algorithm is used.

Acknowledgments

This work was partially supported by a grant of RFBR (project 16-01-00184 A).

References

1. Cartan H. Differential forms. Houghton Mifflin, Boston, 1970.
2. Chaudhuri P. Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of multivariate analysis*, 1991; 39(2): 246269.
3. Godbillon C. Geometrie differentielle et mecanique analytique. Hermann, Paris, 1969.
4. Kotz S, Balakrishnan N, Johnson NL. Continuous Multivariate Distributions, Models and Applications. John Wiley & Sons, 2000; 1.
5. Melkumova LE, Shatskikh SYa. Solving not completely integrable quantile Pfaffian differential equations. *Vestnik SamGU, Natural sciences series*, 2012; 1(3): 2039. [In Russian]
6. Orlova IS, Shatskikh SYa. Pfaffian differential equations for conditional quantiles of multivariate probability distributions. *Vestnik SamGU, Natural sciences series*, 2010; 2(76): 3247. [In Russian]
7. Reinhard H. Equations differentielles: fondements et applications. Dunod, 2 edition, 1989.
8. Shatskikh SYa. A necessary condition of conditional quantile reproducibility of multivariate probability distributions. *Izvestiya RAEN, MMMIU*, 2000; 4(4): 6772. [In Russian]

9. Shatskikh SYa, Knutova EM. Conditional quantile reproducibility of the multivariate Student's distribution. *Izvestiya RAEN, MMMIU*, 1997; 1(1): 3658. [In Russian]
10. Zacks S. *The theory of statistical inference*. John Wiley & Sons, New York, 1971.