October 17, 2012

1 Multivariate Bayesian Quantile Regression with Pólya Tree

1.1 Multivariate Pólya Tree

Due to the difficulty in definition of quantile in multivariate case and diversity of partition methods, there are only few literatures about Pólya tree priors for multivariate data. Paddock (1999, 2002) extended univariate Pólya tree (Lavine 1992, 1994) to multivariate case based on a q-dimensional hypercube. Partitions are constructed through a series of binary recursive perpendicular splits of each axis of the hypercube. Hanson (2006) proposed a q-dimensional location-scale mixture of finite Pólya tree which is a direct generalization of the univariate finite location-scale Pólya tree. Jara (2009) extended the multivariate Pólya tree prior based on Hanson (2006) with an additional parameter: directional orthogonal matrix.

Based on Hanson (2006) and Jara (2009), we briefly introduce multivariate Pólya tree prior as follows: Let $E = \{0, 1\}$, $E^m = \{0, 1\}^m$ be m-fold product of E, and $\pi_j = \left\{B_{\epsilon_{11}\cdots\epsilon_{1j};\cdots;\epsilon_{q1}\cdots\epsilon_{qj}}; \epsilon_{ij} \in E\right\}$ be a level j partition set of Ω such that π_{j+1} are the 2^q finer partitions of π_j.

Definition 1.1.1 (Multivariate Pólya Tree Distribution) *A q-dimensional random probability measure G is said to have a multivariate Pólya tree distribution with parameters $(\Pi, \mathcal{A})$, if there exists nonnegative numbers $\mathcal{A} = \left\{\alpha_{\varepsilon_1;\cdots;\varepsilon_q}; \varepsilon_1, \cdots, \varepsilon_q \in E^j, j =\right.$ (note: ε_i indicates which position the bin takes in level j with respect to i^{th} dimension) and random vectors $\mathcal{Y} = \left\{Y_{\varepsilon_1;\cdots;\varepsilon_q}; \varepsilon_1, \cdots, \varepsilon_q \in E^j, j = 1, \cdots\right\}$, such that the following hold:*

1. *All of the random vectors in $\mathcal{Y}$ are independent,*
2. *For $j = 1, \ldots$ and for all $\varepsilon_1, \cdots, \varepsilon_q \in E^j$, $\boldsymbol{Y}_{\varepsilon_1;\cdots;\varepsilon_q} \sim \text{Dirichlet}\left(\boldsymbol{\alpha}_{\varepsilon_1;\cdots;\varepsilon_q}\right)$, where $\boldsymbol{Y}_{\varepsilon_1;\cdots;\varepsilon_q} = \left\{y_{\varepsilon_1\epsilon_1;\cdots;\varepsilon_q\epsilon_q}; \epsilon_1, \cdots, \epsilon_q \in E\right\}$ and $\boldsymbol{\alpha}_{\varepsilon_1;\cdots;\varepsilon_q} = \left\{\alpha_{\varepsilon_1\epsilon_1;\cdots;\varepsilon_q\epsilon_q}; \epsilon_1, \cdots, \epsilon_q \in E\right\}$,*
3. *For every $j = 1, 2, \cdots$,*
$$G(B_{\epsilon_{11},\cdots,\epsilon_{1j};\cdots;\epsilon_{q1}\cdots\epsilon_{qj}}) = \prod_{l=1}^{j} Y_{\epsilon_{11}\cdots\epsilon_{1l};\cdots;\epsilon_{q1}\cdots\epsilon_{ql}}.$$

Similar to univariate Pólya tree, the canonical way of partition construction is based on reverting cdf of the centering distribution. First, suppose G_0 is a univariate cdf and its corresponding pdf is $g_0(\omega)$. Define $g_0(\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)) = \prod_{i=1}^{q} g_0(\omega_i)$. Denote $\boldsymbol{\theta} = (\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_{q\times q}$ as location-scale parameters, then a family of location-scale baseline measures for multivariate Pólya tree have the following pdf forms $g_{\boldsymbol{\theta}}(\boldsymbol{\omega}) = |\boldsymbol{\Sigma}|^{-1/2} g_0(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\omega} - \boldsymbol{\mu}))$.

For baseline measure $g_0(\boldsymbol{\omega})$, the partition Π_0^j of R^q are obtained from cross-products of corresponding univariate partition sets. Denote

$$B_0(\epsilon_{11}\cdots\epsilon_{1j};\cdots;\epsilon_{q1}\cdots\epsilon_{qj}) = B_0(e_j(k_1)) \times B_0(e_j(k_2)) \times \cdots \times B_0(e_j(k_q)),$$

where $B_0(e_j(k)) = \left(G_0^{-1}((k-1)2^{-j}), G_0^{-1}(k2^{-j})\right)$.

Denote $\boldsymbol{e}_j(\boldsymbol{k}) = e_j(k_1); \cdots; e_j(k_q)$, where $\boldsymbol{k} = (k_1, \ldots, k_q)$, then partitions Π_{θ}^j from location-scale baseline measure family $G_{\boldsymbol{\theta}}$ or $g_{\boldsymbol{\theta}}(\boldsymbol{\omega})$ are defined as

$$B_{\boldsymbol{\theta}}(\boldsymbol{e}_j(\boldsymbol{k})) = \left\{\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{y}; \boldsymbol{y} \in B_0(\boldsymbol{e}_j(\boldsymbol{k}))\right\}.$$

Jara (2009) pointed out that the direction of the sets in Hanson (2006) is complete defined by the decomposition of the covariance matrix, the unique symmetric square root. Instead, he introduced another orthogonal matrix as additional parameter to control the direction of the sets.

Suppose $\boldsymbol{\Sigma} = \boldsymbol{T}'\boldsymbol{T}$, where $\boldsymbol{T}$ is the unique upper triangular Cholesky matrix, then for any orthogonal matrix $\boldsymbol{O}$, let $\boldsymbol{U} = \boldsymbol{O}\boldsymbol{T}$, then $\boldsymbol{U}$ is also a square root of $\boldsymbol{\Sigma}$. Therefore, if we put a prior for $\boldsymbol{O}$ on the space of all $q \times q$ orthogonal matrices, then we have a prior on all possible square roots of $\boldsymbol{\Sigma}$, which control the direction of the partition sets.

The uniqueness in Lemma 1 (Jara 2009) can show it is well defined. In this way, the location-scale transformation induced partition sets $B_{\boldsymbol{\theta}}(\boldsymbol{e}_j(\boldsymbol{k})) = \{\boldsymbol{\mu} + \boldsymbol{T}'\boldsymbol{O}'\boldsymbol{z}; \boldsymbol{z} \in B_0(\boldsymbol{e}_j(\boldsymbol{k}))\}$. The Haar measure (Halmos 1950) provides an easy way to sample orthogonal matrix $\boldsymbol{O}$ "uniformly".

1.2 Multivariate Regression with Pólya Tree

State some motivation

In order to address clustered or correlated data, we propose to model multivariate errors directly, instead of adding random effects. We assume each component of subject's multivariate response can be affected by covariates respectively on its mean and variance, therefore we propose a heterogeneous q-dimensional multivariate regression model:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{B} + (\boldsymbol{X}_i\boldsymbol{\Gamma}) \circ \boldsymbol{E}_i,$$

where $\boldsymbol{Y}_i = [y_{i1}, \dots, y_{iq}]^T$, $\boldsymbol{X}_i = [x_{i1}, \cdots, x_{ip}]$, $\boldsymbol{E}_i = [\epsilon_{i1}, \cdots, \epsilon_{iq}]$,

$$\boldsymbol{B}_{p\times q} = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1q} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pq} \end{bmatrix} \Gamma_{p\times q} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1q} \\ \vdots & \ddots & \vdots \\ \gamma_{p1} & \cdots & \gamma_{pq} \end{bmatrix}$$

so suppose there n subjects and q dimensional responses for each subject:

$$\begin{aligned} \begin{bmatrix} y_{11} & \cdots & y_{1q} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nq} \end{bmatrix}_{n\times q} &= \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}_{n\times p} \begin{bmatrix} \beta_{11} & \cdots & \beta_{1q} \\ \vdots & \ddots & \vdots \\ \beta_{p1} & \cdots & \beta_{pq} \end{bmatrix}_{p\times q} \\ &+ \left(\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1q} \\ \vdots & \ddots & \vdots \\ \gamma_{p1} & \cdots & \gamma_{pq} \end{bmatrix} \right) \circ \begin{bmatrix} \epsilon_{11} & \cdots & \epsilon_{1q} \\ \vdots & \ddots & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nq} \end{bmatrix}_{n\times q} \end{aligned} \tag{1}$$

in which $'\circ'$ is Hadamard product, a.k.a. entrywise product. We assign a multivariate Pólya tree prior on the error:

$$\begin{aligned} \boldsymbol{E}_i = [\epsilon_{i1}, \cdots, \epsilon_{iq}]^T &\overset{\text{i.i.d}}{\sim} G_{\boldsymbol{\theta}} \\ G_{\boldsymbol{\theta}} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{O} &\sim PT\left(\Pi^{\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{O}}, \mathcal{A}\right). \end{aligned}$$

Furthermore, in order not to confound with $\boldsymbol{\beta}$ estimates, we set $\boldsymbol{\mu} = \boldsymbol{0}$ and medians for each component of $G_{\boldsymbol{\theta}}$ are fixed at 0. As to heterogeneity parameters γ_{ij}, for the same reason, we restrict $\gamma_{1j} = 1$ and for all $\boldsymbol{\gamma_{.j}}, \boldsymbol{x}_{i.}$, $\boldsymbol{x_{i.}}\boldsymbol{\gamma_{.j}} > 0$ for all i, j.

Analog to univariate quantile regression with Pólya tree, the posterior τth quantile regression coefficient can be obtained from posterior estimates of

$$\boldsymbol{\beta^{(i)}}(\tau) = \boldsymbol{\beta^{(i)}} + \boldsymbol{\gamma^{(i)}} F^{-1}_{\epsilon^{(i)}}(\tau), \tag{2}$$

where $\boldsymbol{\beta^{(i)}}$ and $\boldsymbol{\gamma^{(i)}}$ is the i^{th} column of $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$, and $F^{-1}_{\epsilon^{(i)}}$ is the inverse marginal cdf of i^{th} component of q-dim error, which can be calculated in the same way from univariate Pólya tree after collapsing multivariate residuals.

1.3 Comparison to Reich

(reich 2010) proposed a flexible Bayesian approach dealing with clustered data based on both conditional and marginal model. He added a random effect term to the flexible Bayesian model in order to address compound symmetric correlation struction. However, the proposed method by us can deal with any correlation struction since Pólya tree can capture the error distribution after 'learning' through enough data observations. More specifically, Reich's method restricts the correlation to be positive and constant across components, while Pólya tree (or an abbr name for our method) works fine for negative correlation, autoregressive model or any other scenarios.

In addition, Reich's approach can only make inference for the quantiles of dependent variables itself, relative to combinatioins of components in a multivariate dependent random vector. For example, quantiles of measurement difference between baseline and couples of weeks are usually of interest in clinical trials, say median$(y_3 - y_0)$, where y_0 and y_3 are observations in baseline and three weeks later respectively. While our method can accommodate the issue by making inference through posterior sampling of $y_3 - y_0$ and using Pólya tree technique to draw posterior quantiles.

1.4 Simulation

The closest approach for multivariate quantile regression is by (Reich 2009), the flexible Bayesian method. Although restricted in compound symmetric correlation structure, it is still competative and shows robustness. The generic function made by Reich is not suitable for the situation where quantile regression coefficients for different components of subject response might differ with each other . Instead Reich's approach assumes the effects of covariates on each components' quantile keep the same. When we do simulations comparing these two methods, in order to adjust the difference, we adapt Reich's function to make the function detect different effects of covariates on quantiles by expanding covariates matrix and the number of corresponding coefficients.

(Reich 2010) considered the model:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{x}_i\boldsymbol{\gamma}\epsilon_i, \tag{3}$$

while for clustered data, the random effect model is adopted:

$$y_{is} = \boldsymbol{x}_{is}\boldsymbol{\beta} + \boldsymbol{x}_{is}\boldsymbol{\gamma}(\alpha_s + \epsilon_{is}), \tag{4}$$

where subscript s indicates s^{th} subject and i indicates i^{th} observation for s^{th} subject. In Reich's scenario, quantile regression coefficients $\boldsymbol{\beta}$ keep the same for all component. Therefore, we modified Reich's function in order to achieve more flexibility in regression coefficients.

Suppose there is only one covariate and there are two observations for each subject. According to (bqrpt) settings, for each subject i, the model is

$$\begin{bmatrix} y_{i1}, & y_{i2} \end{bmatrix} = \begin{bmatrix} 1, & x_i \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} + \left(\begin{bmatrix} 1, & x_i \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \right) \circ \begin{bmatrix} \epsilon_{i1}, & \epsilon_{i2} \end{bmatrix}.$$

For Reich's model, we expand the covariates matrix and regression coefficients,

$$\begin{bmatrix} y_{i1} \\ y_{i2} \end{bmatrix} = \begin{bmatrix} 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_i \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix},$$

where $\boldsymbol{\epsilon}_i = [\epsilon_{i1}, \epsilon_{i2}]$ is a correlated two dimensional error. Evaluation of methods is based on mean squared error:

$$\text{MSE} = \frac{1}{p}\sum_{j=1}^{p}\left(\hat{\beta}_j(\tau) - \beta_j(\tau)\right)^2,$$

where p is the number of covariates except the intercept. $\hat{\beta}_j(\tau)$ is the estimated quantile regression coefficients, and $\beta_j(\tau)$ is the true value. Each component and each quantile is compared between two methods.

1.5 Demonstration

Here we illustrate our approach in several examples. Suppose we have a bunch of bivariate data ($q = 2$) , and there are $n = 100$ observation subjects in each dataset. Details about four datasets are listed below:

M1: $$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \circ \boldsymbol{\epsilon}_{1i}, \boldsymbol{\epsilon_{1i}} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

M2: $$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \circ \boldsymbol{\epsilon}_{2i}, \boldsymbol{\epsilon_{2i}} \sim \text{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right).$$

M3: $$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \circ \boldsymbol{\epsilon}_{3i},$$

$$\boldsymbol{\epsilon_{3i}} \sim 0.5 \times \text{N}\left(\begin{pmatrix} -1.35 \\ 0.28 \end{pmatrix}, \begin{pmatrix} 0.15 & 0.02 \\ 0.02 & 0.04 \end{pmatrix} \right) + 0.5 \times \text{N}\left(\begin{pmatrix} 1.35 \\ 0.28 \end{pmatrix}, \begin{pmatrix} 0.15 & 0.02 \\ 0.02 & 0.04 \end{pmatrix} \right).$$

M4: $\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & -0.5 \\ 0 & 0.5 \end{pmatrix} \circ \boldsymbol{\epsilon}_{3i}$

where ϵ_{3i} are coming from Jara (2009) illustration section. Model 1,2,3 (M1, M2, M3) considered homogeneous model with no correlated normal error, correlated normal error and correlated mixture of bivariate normal error respectively. Model 4 (M4) assigned heterogeneity to mixture of bivariate normal error. Estimates from mean of parameters posterior distributions are shown below:

Table 1: Posterior Estimates for Model 1-4

Model	Term	Estimates	True
M1	$\boldsymbol{B}$	$\begin{pmatrix} 1.05 & 0.97 \\ 1.04 & 0.96 \\ 0.89 & 1.06 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$
	$\boldsymbol{\Gamma}$	$\begin{pmatrix} 1.00 & 1.00 \\ 0.02 & -0.20 \\ -0.04 & 0.14 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$
M2	$\boldsymbol{B}$	$\begin{pmatrix} 0.88 & 1.04 \\ 1.96 & 1.21 \\ 3.03 & 0.86 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$
	$\boldsymbol{\Gamma}$	$\begin{pmatrix} 1.00 & 1.00 \\ 0.09 & -0.02 \\ 0.01 & 0.18 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$
M3	$\boldsymbol{B}$	$\begin{pmatrix} 0.95 & 1.22 \\ 1.02 & 0.97 \\ 0.97 & 1.03 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix}$
	$\boldsymbol{\Gamma}$	$\begin{pmatrix} 1.00 & 1.00 \\ 0.04 & -0.07 \\ 0.02 & 0.00 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$
M4	$\boldsymbol{B}$	$\begin{pmatrix} 0.98 & 1.24 \\ 2.02 & 0.80 \\ 2.94 & 1.17 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$
	$\boldsymbol{\Gamma}$	$\begin{pmatrix} 1.00 & 1.00 \\ 0.03 & -0.55 \\ 0.02 & 0.39 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & -0.5 \\ 0 & 0.5 \end{pmatrix}$