Estimating a Mixture of Dependent Dirichlet Distributions

SDA 2017

Ljubljana, Slovenia

Richard Emilion

University of Orléans, France

June 12, 2017

Outline

Warmest Thanks To Professors S. Cerne, V. Batagelj , N. Kejzar, and all the Slovenian organizing committee

Outline

Warmest Thanks To Professors S. Cerne, V. Batagelj , N. Kejzar, and all the Slovenian organizing committee

- **I - Table of probability vectors**
- **II - Dirichlet Mixtures**

Outline

Warmest Thanks To Professors S. Cerne, V. Batagelj , N. Kejzar, and all the Slovenian organizing committee

- **I - Table of probability vectors**
- **II - Dirichlet Mixtures**
- **III - Dependent Dirichlet Mixtures**

Outline

Warmest Thanks To Professors S. Cerne, V. Batagelj , N. Kejzar, and all the Slovenian organizing committee

- **I - Table of probability vectors**
- **II - Dirichlet Mixtures**
- **III - Dependent Dirichlet Mixtures**
- **IV - Nonparametric: Dirichlet multivariate Kernels**

I - Random distributions / Table of pobability vectors

I.1. Values of a Distribution on a partition

- $[a, b], a \leq b$, interval $\subseteq \mathbb{R}$:
  - summarizes a class of observations/data $\in [a, b]$
  - support of a probability distribution P on $\mathbb{R}$, uniform or not.
  $P([a, b]) = 1$.

I.1. Values of a Distribution on a partition

- $[a, b], a \leq b$, interval $\subseteq \mathbb{R}$:
  - summarizes a class of observations/data $\in [a, b]$
  - support of a probability distribution P on $\mathbb{R}$, uniform or not.
  $P([a, b]) = 1$.
- if $A_1, \ldots A_m$ is a fixed measurable partition of $\mathbb{R}$, then $P(A_1), \ldots P(A_m)$ is a probability vector:

$$P(A_l) \geq 0, \quad \sum_{l=1}^{m} P(A_l) = 1 \tag{1}$$

I.1. Values of a Distribution on a partition

- $[a, b], a \leq b$, interval $\subseteq \mathbb{R}$:
  - summarizes a class of observations/data $\in [a, b]$
  - support of a probability distribution P on $\mathbb{R}$, uniform or not.
  $P([a, b]) = 1$.
- if $A_1, \ldots A_m$ is a fixed measurable partition of $\mathbb{R}$, then $P(A_1), \ldots P(A_m)$ is a probability vector:

$$P(A_l) \geq 0, \quad \sum_{l=1}^{m} P(A_l) = 1 \tag{1}$$

- *Remarks*: (1) also holds if P has a stepwise density function (histogram).
  If P is uniform and $a < b$, then $P(A_l) = \frac{|A_l \cap [a,b]|}{b-a}$, $|\ |$ denoting the Lebesgue measure.

I.2. Sample of probability vectors

- $(\Omega, \mathcal{F}, \mathbb{P})$: a probability space.
- $[a_i, b_i], a_i \leq b_i, \; i = 1, \ldots, n$, a sample of intervals.
- $P_i, \; i = 1, \ldots, n$, a sample of distributions: outcomes of a random distribution $P : \Omega \longrightarrow M_1(\mathbb{V})$
  where $M_1(V)$ denotes the space of probabilty measures on $\mathbb{V}$ a closed subset of $\mathbb{R}$, endowed with the weak topology.
- if $A_1, \ldots A_m$ is a fixed measurable partition of $\mathbb{V}$, then $P_i(A_1), \ldots P_i(A_m)$ is a sample of probability vectors.

I.3. Table of probability vectors

- More generally: p dependent random distributions $P_{.j} : \Omega \longrightarrow M_1(\mathbb{V}_j)$, $j = 1, \ldots, p$
- Outcomes $P_{i,j}$, $i = 1, \ldots, n$
- if $A_1, \ldots A_{m_j}$ is a fixed measurable partition of $\mathbb{V}_j$, then $P_{i,j}(A_1), \ldots P_i(A_{m_j})$ is a table of probability vectors.

I.3. Table of probability vectors

- More generally: p dependent random distributions $P_{.j} : \Omega \longrightarrow M_1(\mathbb{V}_j),\ j = 1, \ldots, p$
- Outcomes $P_{i,j},\ i = 1, \ldots, n$
- if $A_1, \ldots A_{m_j}$ is a fixed measurable partition of $\mathbb{V}_j$, then $P_{i,j}(A_1), \ldots P_i(A_{m_j})$ is a table of probability vectors.
- Problem: Which model to fit ? Mixture model ?

II - $p = 1$, Mixtures of Dirichlet Distributions

II.1 $p=1$, Dirichlet Distribution (DD)

- Let $\alpha_l \geq 0$ for $l = 1,\ldots,m_1$, and $X_l \overset{ind}{\sim} \Gamma(\alpha_l, 1)$, i.e. with density $x^{\alpha_l - 1}e^{-x}1_{\mathbb{R}_+}(x)$, then the Dirichlet Distribution (DD), $Dir(\alpha_1,\ldots,\alpha_{m_1})$, is the distribution of the random probability vector $(\frac{X_1}{X_1+\ldots+X_{m_1}},\ldots,\frac{X_{m_1}}{X_1+\ldots+X_{m_1}})$.
- No density, but the density of

$$Y = (Y_1,\ldots,Y_{m_1-1}) = (\frac{X_1}{X_1+\ldots+X_{m_1}},\ldots,\frac{X_{m_1-1}}{X_1+\ldots+X_{m_1}}) \text{ is}$$

$$\frac{\Gamma(\alpha_1+\ldots+\alpha_{m_1})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{m_1})} y_1^{\alpha_1-1}\ldots y_{m_1-1}^{\alpha_{m_1-1}-1}(1-\sum_{i=1}^{m_1-1} y_i)^{\alpha_{m_1}-1} I_{T_{m_1-1}}(y)$$

with
$T_{m_1-1} = \{y = (y_1,\ldots,y_{m_1-1}) \in \mathbb{R}_+^{m_1-1} : \sum_{l=1}^{m_1-1} y_l \leq 1\}$.
Y is independent of
$Z_1 = X_1+\ldots+X_{m_1} \sim \Gamma(\alpha_1+\ldots+\alpha_{m_1}, 1)$.

II.2 $p = 1$, DD Mixtures

- $\underline{\alpha} = (\alpha_1, \ldots, \alpha_l, \ldots, \alpha_{m_1})$ estim.: MLE (Mika 1986, 2000)

II.2 $p = 1$, DD Mixtures

- $\underline{\alpha} = (\alpha_1, \ldots, \alpha_l, \ldots, \alpha_{m_1})$ estim.: MLE (Mika 1986, 2000)
- DD belongs to the exponential family: any DD Mixture $\sum_{h=1}^{K} q_h DD(\underline{\alpha}_h)$, can be estimated by the EM algorithm

II.2 $p = 1$, DD Mixtures

- $\underline{\alpha} = (\alpha_1, \ldots, \alpha_l, \ldots, \alpha_{m_1})$ estim.: MLE (Mika 1986, 2000)
- DD belongs to the exponential family: any DD Mixture $\sum_{h=1}^{K} q_h DD(\underline{\alpha}_h)$, can be estimated by the EM algorithm
- Unobserved (latent) class variable
  $C : \Omega \longrightarrow \{1, \ldots, K\} \; : \; \mathbb{P}(C = h) = q_h$
  Observed variable X : $\mathbb{P}_{X|C=h} = DD(\underline{\alpha}_h)$ so that

$$\mathbb{P}_X = \sum_{h=1}^{K} q_h DD(\underline{\alpha}_h)$$

while the degree of which x belongs to *fuzzy* class h is:

$$t_{x,h} = P_{C=h|X=x} = \frac{q_h DD(\underline{\alpha}_h)(x)}{\sum_{r=1}^{K} q_r DD(\underline{\alpha}_r)(x)}$$

II.3 $p = 1$, DD Mixture, Log Likelihood

- Complete variable (X, C) likelihood for a 1-sample:

$$L(x, c) = DD(\underline{\alpha}_c)(x)q_c = \prod_{h=1}^{K} (DD(\underline{\alpha}_h)(x)q_h)^{1_{c=h}}$$

- Complete variable (X, C) Likelihood for a n-sample $(\underline{x}, \underline{c}) = (x_i, c_i)_{i=1,\dots,n}$:

$$L(\underline{x}, \underline{c}) = \prod_{i=1}^{n} \prod_{h=1}^{K} (DD(\underline{\alpha}_h)(x_i)q_h)^{1_{c_i=h}}$$

and Log Likelihood is

$$LL(\underline{x}, \underline{c}, \alpha, q) = \sum_{i=1}^{n} \sum_{h=1}^{K} 1_{c_i=h}(\log(DD(\underline{\alpha}_h)(x_i)) + \log(q_h))$$

II.4 $p = 1$, E(Log Likelihood) for a DD Mixture 11 / 24

- E(xpectation) part computing in EM algorithm:
$E_{\mathbb{P}_{C|X,\alpha_{old},q_{old}}}(LL(\underline{X}, \underline{C}, \alpha, q)) =$
$\sum_{i=1}^{n} \sum_{h=1}^{K} t_{x_i,h,\alpha_{old},q_{old}}(\log(DD(\underline{\alpha}_h)(x_i)) + \log(q_h))$

II.4 $p = 1$, E(Log Likelihood) for a DD Mixture 11 / 24

- E(xpectation) part computing in EM algorithm:
  $E_{\mathbb{P}_{C|X,\alpha_{old},q_{old}}}(LL(\underline{X}, \underline{C}, \alpha, q)) =$
  $\sum_{i=1}^{n} \sum_{h=1}^{K} t_{x_i,h,\alpha_{old},q_{old}}(\log(DD(\underline{\alpha}_h)(x_i)) + \log(q_h))$
- $\log(DD(\underline{\alpha}_h)(x_i)) =$
  $= \log\Gamma(\sum_{l=1}^{m_1} \alpha_{h,l}) - \sum_{l=1}^{m_1} \log\Gamma(\alpha_{h,l}) + \sum_{l=1}^{m_1}(\alpha_{h,l} - 1)\log(x_{i,l})$.

II.4 $p = 1$, E(Log Likelihood) for a DD Mixture 11 / 24

- E(xpectation) part computing in EM algorithm:
  $E_{\mathbb{P}_{C|X,\alpha_{old},q_{old}}}(LL(\underline{X}, \underline{C}, \alpha, q)) =$
  $\sum_{i=1}^{n} \sum_{h=1}^{K} t_{x_i,h,\alpha_{old},q_{old}}(\log(DD(\underline{\alpha}_h)(x_i)) + \log(q_h))$
- $\log(DD(\underline{\alpha}_h)(x_i)) =$
  $= \log\Gamma(\sum_{l=1}^{m_1} \alpha_{h,l}) - \sum_{l=1}^{m_1} \log\Gamma(\alpha_{h,l}) + \sum_{l=1}^{m_1}(\alpha_{h,l} - 1)\log(x_{i,l})$.
- M part of EM: new α maximizes the above expectation.
  Derivation: m_1 equations ($l = 1, \ldots, m_1$) for component h:
  $\sum_{i=1}^{n} t_{x_i,h,\alpha_{old},q_{old}}(digamma(\sum_{h=1}^{m_1} \alpha_{h,l}) - digamma(\alpha_{h,l}) + \log(x_{i,l})) = 0$

II.4 $p = 1$, E(Log Likelihood) for a DD Mixture 11 / 24

- E(xpectation) part computing in EM algorithm:
  $E_{\mathbb{P}_{C|X,\alpha_{old},q_{old}}}(LL(\underline{X}, \underline{C}, \alpha, q)) =$
  $\sum_{i=1}^{n} \sum_{h=1}^{K} t_{x_i,h,\alpha_{old},q_{old}}(\log(DD(\underline{\alpha}_h)(x_i)) + \log(q_h))$
- $\log(DD(\underline{\alpha}_h)(x_i)) =$
  $= \log\Gamma(\sum_{l=1}^{m_1} \alpha_{h,l}) - \sum_{l=1}^{m_1} \log\Gamma(\alpha_{h,l}) + \sum_{l=1}^{m_1}(\alpha_{h,l} - 1)\log(x_{i,l})$.
- M part of EM: new α maximizes the above expectation.
  Derivation: m_1 equations ($l = 1, \ldots, m_1$) for component h:
  $\sum_{i=1}^{n} t_{x_i,h,\alpha_{old},q_{old}}(digamma(\sum_{h=1}^{m_1} \alpha_{h,l}) - digamma(\alpha_{h,l}) + \log(x_{i,l})) = 0$
- R Implementation with Bang Xia (PhD Student): Package BB, dfsane function.

II.5 $p = 1$, EM, SEM, CEM, DC

- Bang Xia, R.E., E. Diday, H. Wang, joint paper
- Mixture Estimation. EM-like algorithms: fuzzy classes. DC (Dynamical clustering): crisp classes

II.5 $p = 1$, EM, SEM, CEM, DC

- Bang Xia, R.E., E. Diday, H. Wang, joint paper
- Mixture Estimation. EM-like algorithms: fuzzy classes. DC (Dynamical clustering): crisp classes
- Fuzzy class h, fuzzy mean histogram $\frac{\sum_{i=1}^{n} t_{x_i,h} x_i}{\sum_{i=1}^{n} t_{x_i,h}}$

II.5 $p = 1$, EM, SEM, CEM, DC

- Bang Xia, R.E., E. Diday, H. Wang, joint paper
- Mixture Estimation. EM-like algorithms: fuzzy classes. DC (Dynamical clustering): crisp classes
- Fuzzy class h, fuzzy mean histogram $\frac{\sum_{i=1}^{n} t_{x_i,h} x_i}{\sum_{i=1}^{n} t_{x_i,h}}$
- Define an 'Explanatory' quality criterion of a classification algorithm: Beetween class variance w.r.t. a choosen distance. Compare all these algorithms: Likelihood, Explanatory quality.

II.6 $p = 1$, EM Consistency, npEM

- A consistency result ($m_1 \longrightarrow +\infty$) using Dirichlet processes, was proved for S.E.M.: R.E. (2014)

II.6 $p = 1$, EM Consistency, npEM

- A consistency result ($m_1 \longrightarrow +\infty$) using Dirichlet processes, was proved for S.E.M.: R.E. (2014)
- nonparametric case: Dirichlet Kernel, npEM

III - $p \geq 2$, Mixtures of Dependent Dirichlet Distributions

III.1 $p \geq 2$, Dependent DD model 1

- Finite mixture with independence conditionnally to a class variable C taking values $h = 1, \ldots K$ with probability q_h, resp.

$$\sum_{h=1}^{K} q_h DD_{m_1,h} \otimes DD_{m_2,h} \otimes \ldots \otimes DD_{m_p,h}$$

where $DD_{m_j,h}$ denotes a DD in dimension m_j with parameters depending on h

III.1 $p \geq 2$, Dependent DD model 1

- Finite mixture with independence conditionnally to a class variable C taking values $h = 1, \dots K$ with probability q_h, resp.

$$\sum_{h=1}^{K} q_h DD_{m_1,h} \otimes DD_{m_2,h} \otimes \dots \otimes DD_{m_p,h}$$

  where $DD_{m_j,h}$ denotes a DD in dimension m_j with parameters depending on h

- Estimation using EM-like algorithm

III.1 $p \geq 2$, Dependent DD model 1

- Finite mixture with independence conditionnally to a class variable C taking values $h = 1, \ldots K$ with probability q_h, resp.

$$\sum_{h=1}^{K} q_h DD_{m_1,h} \otimes DD_{m_2,h} \otimes \ldots \otimes DD_{m_p,h}$$

  where $DD_{m_j,h}$ denotes a DD in dimension m_j with parameters depending on h

- Estimation using EM-like algorithm
- Independence conditonnally to C but there is no indepedence of the p histogram variables

III.2 $p \geq 2$, Dependent DD model 2

- Led d large enough ($d > \max(m_1, m_2, \ldots, m_p)$), let $G = (G_1, \ldots, G_d)^t : G_j \overset{ind}{\sim} \Gamma(\alpha_j, 1)$

III.2 $p \geq 2$, Dependent DD model 2

- Led d large enough ($d > \max(m_1, m_2, \ldots, m_p)$), let $G = (G_1, \ldots, G_d)^t : G_j \overset{ind}{\sim} \Gamma(\alpha_j, 1)$
- Choose m_1 integers in $\{1, \ldots, d\}$, divide each of the m_1 corresp. G_j's by their sum: Dirichlet vector in dim m_1. Proceed similarly with m_2, ..., with m_p

III.2 $p \geq 2$, Dependent DD model 2

- Led d large enough ($d > \max(m_1, m_2, \ldots, m_p)$), let $G = (G_1, \ldots, G_d)^t : G_j \overset{ind}{\sim} \Gamma(\alpha_j, 1)$
- Choose m_1 integers in $\{1, \ldots, d\}$, divide each of the m_1 corresp. G_j's by their sum: Dirichlet vector in dim m_1.
  Proceed similarly with m_2, . . . , with m_p
  If some integers in these choices are common: p Dependent Dirichlet vectors.

III.2 $p \geq 2$, Dependent DD model 2

- Led d large enough ($d > \max(m_1, m_2, \ldots, m_p)$), let $G = (G_1, \ldots, G_d)^t : G_j \overset{ind}{\sim} \Gamma(\alpha_j, 1)$
- Choose m_1 integers in $\{1, \ldots, d\}$, divide each of the m_1 corresp. G_j's by their sum: Dirichlet vector in dim m_1.
  Proceed similarly with m_2, ..., with m_p
  If some integers in these choices are common: p Dependent Dirichlet vectors.
- Example $d = 3, p = 2 = m_1 = m_2$. Choosing 1, 2 and then 2, 3 we get 2 dependent Dirichlet vectors
  $(Y_1 = \frac{X_1}{X_1+X_2}, Y_2 = \frac{X_2}{X_1+X_2})$ and $(Y_3 = \frac{X_2}{X_2+X_3}, Y_4 = \frac{X_3}{X_2+X_3})$.
  The joint density of (Y_1, Y_3) is a continous mixture:

$$f_{(Y_1,Y_3)}(y_1, y_3) = \int_{\mathbb{R}_+} f_{X_1}\left(\frac{y_1 x_2}{1-y_1}\right) f_{X_3}\left(\frac{y_2 x_2}{1-y_2}\right) f_{X_2}(x_2) dx_2$$

III.3 $p \geq 2$, Dependent DD model 3 /a

- Let $X \sim \mathcal{N}(0,1)$ with cdf $\Phi : \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$
  Since $\Phi(X) \sim Uniform(0,1)$, if F_α denotes the cdf of $\Gamma(\alpha, 1)$, and $g_\alpha = F_\alpha^{-1} \circ \Phi$, then $g_\alpha(X) \sim \Gamma(\alpha, 1)$

III.3 $p \geq 2$, Dependent DD model 3 /a

- Let $X \sim \mathcal{N}(0,1)$ with cdf $\Phi : \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$
  Since $\Phi(X) \sim Uniform(0,1)$, if F_α denotes the cdf of $\Gamma(\alpha, 1)$, and $g_\alpha = F_\alpha^{-1} \circ \Phi$, then $g_\alpha(X) \sim \Gamma(\alpha, 1)$
- Let $r = m_1 + m_2 + \ldots + m_p$ and Σ be a $r \times r$ symmetric positive definitive matrix having p Identity matrices of size m_j, respectively, as diagonal blocks.
  For example if $p = 2, m_1 = 2, m_2 = 3$, then $r = 5$ and

$$\Sigma = \begin{pmatrix} 1 & 0 & a & b & c \\ 0 & 1 & d & e & f \\ a & d & 1 & 0 & 0 \\ b & e & 0 & 1 & 0 \\ c & f & 0 & 0 & 1 \end{pmatrix}$$

III.4 $p \geq 2$, Dependent DD model 3 /b

- Starting from a Gaussian (starting from Dependent Gammas is another option).

III.4 $p \geq 2$, Dependent DD model 3 /b

- Starting from a Gaussian (starting from Dependent Gammas is another option).
- Let $G = (G_1, \ldots, G_r) \sim Gauss(0, \Sigma)$ be a Gaussian vector in dimension r. Due to the pattern of Σ, $G_1, \ldots G_{m_1}$ are i.i.d. $\mathcal{N}(0,1)$, the same for $G_{m_1+1}, \ldots G_{m_1+m_2}$ and so on, up to $G_{m_1+\ldots+m_{p-1}+1}, \ldots, G_r$. However these p groups of vectors are dependent each other if the correlation parameters are $\neq 0$.

III.4 $p \geq 2$, Dependent DD model 3 /b

- Starting from a Gaussian (starting from Dependent Gammas is another option).
- Let $G = (G_1, \ldots, G_r) \sim Gauss(0, \Sigma)$ be a Gaussian vector in dimension r. Due to the pattern of Σ, $G_1, \ldots G_{m_1}$ are i.i.d. $\mathcal{N}(0,1)$, the same for $G_{m_1+1}, \ldots G_{m_1+m_2}$ and so on, up to $G_{m_1+\ldots+m_{p-1}+1}, \ldots, G_r$. However these p groups of vectors are dependent each other if the correlation parameters are $\neq 0$.
- Let $\alpha_j \geq 0, j = 1, \ldots, r$ be some parameters and let $X_j = g_{\alpha_j}(G_j)$ so that $X_1, \ldots X_{m_1}$ are independent Gammas with parameters $(\alpha_1, 1), \ldots (\alpha_{m_1}, 1)$, respectively. The same for $X_{m_1+1}, \ldots X_{m_1+m_2}$ and so on, up to $X_{m_1+\ldots+m_{p-1}+1}, \ldots, X_r$. But these groups of Gamma vectors are dependent each other.

III.5 $p \geq 2$, Dependent DD model 3 /c

- The joint density of the X_j's is derived from the Gaussian density:

$$\frac{1}{\sqrt{(2\pi)^r|\Sigma|}} \exp^{-\frac{1}{2}v^t\Sigma^{-1}v} \prod_{j=1}^{r} (g_{\alpha_j}^{-1})'(x_j)$$

where $v = (g_{\alpha_1}^{-1}(x_1), \ldots, g_{\alpha_r}^{-1}(x_r))^t$ and

$$(g_{\alpha_j}^{-1})'(x_j) = \frac{\sqrt{2\pi}}{\Gamma(\alpha_j)} \exp^{\frac{[\Phi^{-1}(F_{\alpha_j}(x_j))]^2}{2} - x_j} x_j^{\alpha_j - 1}$$

Diviging each of these Gammas by the sum of their respective group, we get p Dirichlet vectors which are dependent.

III.6 $p \geq 2$, Dependent DD model 3 /d

Consider the following bijections

- $Q_{1:m_1} : \mathbb{R}_+^{m_1} \longrightarrow T_{m_1 - 1} \times \mathbb{R}_+$ used when $p = 1$
  $Q_{1:m_1}(x_1, \ldots, x_{m_1}) = (y_1, \ldots, y_{m_1 - 1}, z_1)$ with

$$y_1 = \frac{x_1}{x_1 + \ldots + x_{m_1}}, \ldots, y_{m_1 - 1} = \frac{x_{m_1 - 1}}{x_1 + \ldots + x_{m_1}}, z_1 = x_1 + \ldots + x_{m_1}$$

  Jacobian of $Q_{1:m_1}^{-1}$ is $z_1^{m_1 - 1}$

III.6 $p \geq 2$, Dependent DD model 3 /d

Consider the following bijections

- $Q_{1:m_1} : \mathbb{R}_+^{m_1} \longrightarrow T_{m_1-1} \times \mathbb{R}_+$ used when $p = 1$
  $Q_{1:m_1}(x_1, \ldots, x_{m_1}) = (y_1, \ldots, y_{m_1-1}, z_1)$ with
  $$y_1 = \frac{x_1}{x_1 + \ldots + x_{m_1}}, \ldots, y_{m_1-1} = \frac{x_{m_1-1}}{x_1 + \ldots + x_{m_1}}, z_1 = x_1 + \ldots + x_{m_1}$$
  Jacobian of $Q_{1:m_1}^{-1}$ is $z_1^{m_1-1}$
- $Q_{m_1+1:m_1+m_2} : \mathbb{R}_+^{m_2} \longrightarrow T_{m_2-1} \times \mathbb{R}_+$
  $Q_{m_1+1:m_1+m_2}(x_{m_1+1}, \ldots, x_{m_1+m_2}) = (y_{m_1+1}, \ldots, y_{m_1+m_2-1}, z_2)$
  with
  $$y_{m_1+1} = \frac{x_{m_1+1}}{x_{m_1+1} + \ldots + x_{m_1+m_2}}, \ldots, y_{m_1+m_2-1} = \frac{x_{m_1+m_2-1}}{x_{m_1+1} + \ldots + x_{m_1+}}$$
  $z_2 = x_{m_1+1} + \ldots + x_{m_1+m_2}$. Jacobian of $Q_{m_1+1:m_1+m_2}^{-1}$ is $z_2^{m_1-1}$

III.6 $p \geq 2$, Dependent DD model 3 /d

Consider the following bijections

- $Q_{1:m_1} : \mathbb{R}_+^{m_1} \longrightarrow T_{m_1-1} \times \mathbb{R}_+$ used when $p = 1$
  $Q_{1:m_1}(x_1, \ldots, x_{m_1}) = (y_1, \ldots, y_{m_1-1}, z_1)$ with
  $$y_1 = \frac{x_1}{x_1 + \ldots + x_{m_1}}, \ldots, y_{m_1-1} = \frac{x_{m_1-1}}{x_1 + \ldots + x_{m_1}}, z_1 = x_1 + \ldots + x_{m_1}$$
  Jacobian of $Q_{1:m_1}^{-1}$ is $z_1^{m_1-1}$
- $Q_{m_1+1:m_1+m_2} : \mathbb{R}_+^{m_2} \longrightarrow T_{m_2-1} \times \mathbb{R}_+$
  $Q_{m_1+1:m_1+m_2}(x_{m_1+1}, \ldots, x_{m_1+m_2}) = (y_{m_1+1}, \ldots, y_{m_1+m_2-1}, z_2)$
  with
  $$y_{m_1+1} = \frac{x_{m_1+1}}{x_{m_1+1} + \ldots + x_{m_1+m_2}}, \ldots, y_{m_1+m_2-1} = \frac{x_{m_1+m_2-1}}{x_{m_1+1} + \ldots + x_{m_1+}}$$
  $z_2 = x_{m_1+1} + \ldots + x_{m_1+m_2}$. Jacobian of $Q_{m_1+1:m_1+m_2}^{-1}$ is $z_2^{m_1-1}$
- $\ldots Q_{m_1+\ldots+m_{p-1}+1:r} \cdots$

III.7 $p \geq 2$, Dependent DD model 3 /e

Applying these bijections to the X_j's, we get the density of the random vector $(Y_1, \ldots, Y_{m_1-1}, Z_1, Y_{m_1+1}, \ldots, Y_{m_1+m_2-1}, Z_2, \ldots, Y_{m_1+\ldots+m_{p-1}+1}, \ldots, Y_{r-1}, Z_p)$:

$$\frac{1}{\sqrt{(2\pi)^r|\Sigma|}} \exp^{-\frac{1}{2}w^t\Sigma^{-1}w} \prod_{j=1}^{r} (g_{\alpha_j}^{-1})'(u_j) z_1^{m_1-1} z_2^{m_2-1} \ldots z_p^{m_p-1}$$

where $w = (g_{\alpha_1}^{-1}(u_1), \ldots, g_{\alpha_j}^{-1}(u_j), \ldots, g_{\alpha_j}^{-1}(u_r))^t$ and
$u_j = y_j z_1$ if $1 \leq j < m_1$, $u_{m_1} = z_1(1 - y_1 - \ldots - y_{m_1-1})$
$u_j = y_j z_2$ if $m_1 + 1 \leq j < m_1 + m_2$,
$u_{m_1+m_2} = z_2(1 - y_{m_1+1} - \ldots - y_{m_1+m_2-1}) \ldots, \ldots$
$u_j = y_j z_p$ if $m_1 + \ldots + m_{p-1} + 1 \leq j < r = m_1 + \ldots + m_p$,
$u_r = z_p(1 - y_{m_1+\ldots+m_{p-1}+1} - \ldots - y_{r-1})$

III.8 $p \geq 2$, Dependent DD model 2 /f

- This yields the joint density of the Dependent Dirichlet $(Y_1, \ldots, Y_{m_1-1}, Y_{m_1+1}, \ldots, Y_{m_1+m_2-1}, \ldots, Y_{m_1+\ldots+m_{p-1}+1}, \ldots, Y_{r-1}$ by integrating out the above density w.r.t $z_1, z_2, \ldots, z_p$

III.8 $p \geq 2$, Dependent DD model 2 /f

- This yields the joint density of the Dependent Dirichlet $(Y_1, \ldots, Y_{m_1-1}, Y_{m_1+1}, \ldots, Y_{m_1+m_2-1}, \ldots, Y_{m_1+\ldots+m_{p-1}+1}, \ldots, Y_{r-1}$ by integrating out the above density w.r.t $z_1, z_2, \ldots, z_p$
- This integral can be estimated by simulation since (Y, Z) can be easily simulated

III.8 $p \geq 2$, Dependent DD model 2 /f

- This yields the joint density of the Dependent Dirichlet $(Y_1, \ldots, Y_{m_1-1}, Y_{m_1+1}, \ldots, Y_{m_1+m_2-1}, \ldots, Y_{m_1+\ldots+m_{p-1}+1}, \ldots, Y_{r-1}$ by integrating out the above density w.r.t $z_1, z_2, \ldots, z_p$
- This integral can be estimated by simulation since (Y, Z) can be easily simulated
- Estimation of K and of the α_j's: MCEM algorithm ? Formal calculus software (work in progress with A. Roy & G. Levey) ?

IV.1 Dirichlet Kernels

- $p = 1$. Dirichlet density $Dd(\alpha_1, \dots, \alpha_{m_1})$ = multivariate kernel when the α_j's are chosen (e.g. $\alpha_j = 1, \forall j$). Observations: $o_1, \dots, o_n \in T_{m_1 - 1}$, density estimation:

$$K_H(x) = |H|^{-\frac{1}{2}} \sum_{i=1} nDd(H^{-\frac{1}{2}}(x - o_i)), x \in T_{m_1 - 1}$$

$H : (m_1 - 1) \times (m_1 - 1)$ symmetric postive definitive bandwith matrix to be determined.

IV.1 Dirichlet Kernels

- $p = 1$. Dirichlet density $Dd(\alpha_1, \ldots, \alpha_{m_1})$ = multivariate kernel when the α_j's are chosen (e.g. $\alpha_j = 1, \forall j$). Observations: $o_1, \ldots, o_n \in T_{m_1-1}$, density estimation:

$$K_H(x) = |H|^{-\frac{1}{2}} \sum_{i=1} nDd(H^{-\frac{1}{2}}(x - o_i)), x \in T_{m_1-1}$$

  $H : (m_1 - 1) \times (m_1 - 1)$ symmetric postive definitive bandwith matrix to be determined.

- $p = 1$. Use Dirichlet kernels to estimate a mixture of R densities in T_{m_1-1}: bandwith matrices $H_c, c = 1, \ldots, R$ depending on component c, on the algorithm iteration.

IV.1 Dirichlet Kernels

- $p = 1$. Dirichlet density $Dd(\alpha_1, \ldots, \alpha_{m_1})$ = multivariate kernel when the α_j's are chosen (e.g. $\alpha_j = 1, \forall j$). Observations: $o_1, \ldots, o_n \in T_{m_1 - 1}$, density estimation:

$$K_H(x) = |H|^{-\frac{1}{2}} \sum_{i=1} nDd(H^{-\frac{1}{2}}(x - o_i)), x \in T_{m_1-1}$$

  $H : (m_1 - 1) \times (m_1 - 1)$ symmetric postive definitive bandwith matrix to be determined.

- $p = 1$. Use Dirichlet kernels to estimate a mixture of R densities in $T_{m_1 - 1}$: bandwith matrices $H_c, c = 1, \ldots, R$ depending on component c, on the algorithm iteration.
- $p \geq 2$. Each component which is the product of its p marginals can be estimated as a product of p Dirichlet kernels.

IV.2 Dependent Dirichlet Kernels

- $p \geq 2$. Method 1. Use p Dd kernels and consider that their product estimate each component of a mixture as in mvnpEM, Chauveau-Hoang, *CSDA*, 2016

IV.2 Dependent Dirichlet Kernels

- $p \geq 2$. Method 1. Use p Dd kernels and consider that their product estimate each component of a mixture as in mvnpEM, Chauveau-Hoang, *CSDA*, 2016
- $p \geq 2$. Method 2. Fix the Gaussian covariance matrix Σ and the Gammas parameters $\alpha_1, \ldots \alpha_r$ and consider the Dependent Dirichlet density as a multivariate kernel and proceed to a multivariate density estimation based on that Kernel.

IV.2 Dependent Dirichlet Kernels

- $p \geq 2$. Method 1. Use p Dd kernels and consider that their product estimate each component of a mixture as in mvnpEM, Chauveau-Hoang, *CSDA*, 2016
- $p \geq 2$. Method 2. Fix the Gaussian covariance matrix Σ and the Gammas parameters $\alpha_1, \ldots \alpha_r$ and consider the Dependent Dirichlet density as a multivariate kernel and proceed to a multivariate density estimation based on that Kernel.
- $p \geq 2$. Method 3. Assuming dependencies as in II.5, use the multivariate density with choosen parameters as a kernel and proceed as above.