Appendix

Details on Bound (24)

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}_{P_0}\left[\left(f(P_0)-\beta^T Z(\widehat{P}_0)\right)^2\right] \\
&= \frac{1}{2}\psi^T\Sigma\psi + \mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)^T\right]\psi + \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] \\
&\quad - \frac{1}{2}\psi^T\Sigma\psi - \mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)^T\right]\Sigma^+\Sigma\psi \\
&\quad - \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)^T\right]\Sigma^+\mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)^T\left(\psi - \Sigma^+\Sigma\psi\right)\right],
\end{aligned}
$$

since Σ^+ is PSD. Let $\Sigma = USU^{-1}$ be the eigen-decomposition of Σ; i.e. S is diagonal matrix of decreasing eigenvalues and U is a real unitary matrix and $U^{-1} = U^T$. Then, $\Sigma^+ = US^+U^{-1}$, where S^+ is the diagonal matrix where $(S^+)_{ii} = 1/(S)_{ii}$ if $(S)_{ii} \neq 0$ and $(S^+)_{ii} = 0$ if $(S)_{ii} = 0$. Furthermore, let $r = \text{rank}(\Sigma)$, and I_r be the diagonal matrix with $(I_r)_{ii} = 1$ for $i \leq r$ and $(I_r)_{ii} = 0$ for $i > r$. Hence:

$$
\begin{aligned}
\|\Sigma^+\Sigma\psi\|_2^2 &= \psi^T\Sigma\Sigma^+\Sigma^+\Sigma\psi \\
&= \psi^T USU^{-1}US^+U^{-1}US^+U^{-1}USU^{-1}\psi \\
&= \psi^T UI_rU^{-1}\psi \\
&\leq \psi^T UIU^{-1}\psi \\
&\leq \|\psi\|_2^2.
\end{aligned}
$$

Furthermore,

$$\|\psi\|_2 \leq \sum_{i=1}^{\infty}|\boldsymbol{\theta}_i|\|Z(G_i)\|_2 \leq \sqrt{2}B.$$

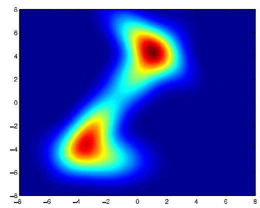
Hence,

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}_{P_0}\left[\left(f(P_0)-\beta^T Z(\widehat{P}_0)\right)^2\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \mathbb{E}_{P_0}\left[\varsigma_0 Z(\widehat{P}_0)^T\left(\psi - \Sigma^+\Sigma\psi\right)\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \mathbb{E}_{P_0}\left[|\varsigma_0||Z(\widehat{P}_0)^T\left(\psi - \Sigma^+\Sigma\psi\right)|\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \mathbb{E}_{P_0}\left[|\varsigma_0|\|Z(\widehat{P}_0)\|_2\|\psi - \Sigma^+\Sigma\psi\|_2\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \sqrt{2}(\|\psi\|_2 + \|\Sigma^+\Sigma\psi\|_2)\mathbb{E}_{P_0}\left[|\varsigma_0|\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \sqrt{2}(\|\psi\|_2 + \|\psi\|_2)\mathbb{E}_{P_0}\left[|\varsigma_0|\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + \sqrt{2}(2\sqrt{2}B)\mathbb{E}_{P_0}\left[|\varsigma_0|\right] \\
&\leq \frac{1}{2}\mathbb{E}_{P_0}\left[\varsigma_0^2\right] + 4B\sqrt{\mathbb{E}_{P_0}\left[\varsigma_0^2\right]},
\end{aligned}
$$

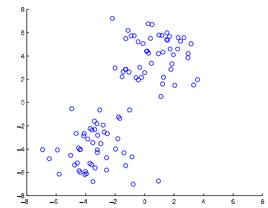
where the last line follows from Jensen's inequality.

GMM Figures

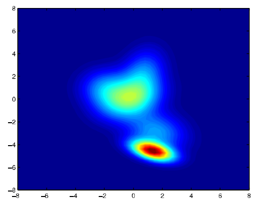
See Figure 5.



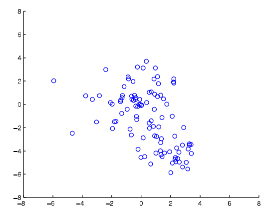
(a) $k = 4$ True Density



(b) $k = 4$, $n = 100$ Points



(c) $k = 5$ True Density



(d) $k = 5$, $n = 100$ Points

Figure 5: Typical GMMs generated in our datasets as well as their corresponding samples. One can see that it would be hard for even a human to predict the true number of components, yet the Double-Basis estimator does a good job.

Density Estimation Details

Let $M_t = \{\alpha \;:\; \kappa_\alpha(\nu,\gamma) \leq t\} = \{\alpha_1,\dots,\alpha_S\}$. First note that:

$$
\begin{aligned}
&\mathbb{E}\left[\|p_i - \tilde{p}_i\|_2^2\right] \\
=&\mathbb{E}\left[\left\|\sum_{\alpha\in\mathbb{Z}} a_\alpha(P_i)\varphi_\alpha - \sum_{\alpha\in M_t} a_\alpha(\widehat{P}_i)\varphi_\alpha\right\|_2^2\right] \\
=&\mathbb{E}\Bigg[\int_{\Lambda^l}\Bigg(\sum_{\alpha\in M_t}(a_\alpha(P_i) - a_\alpha(\widehat{P}_i))\varphi_\alpha(x) \\
&\qquad + \sum_{\alpha\in M_t^c} a_\alpha(P_i)\varphi_\alpha(x)\Bigg)^2 \mathrm{d}x\Bigg] \\
=&\mathbb{E}\Bigg[\int_{\Lambda^l}\sum_{\alpha\in M_t}\sum_{\rho\in M_t}(a_\alpha(P_i) - a_\alpha(\widehat{P}_i))(a_\rho(P_i) - a_\rho(\widehat{P}_i)) \\
&\qquad \varphi_\alpha(x)\varphi_\rho(x)\mathrm{d}x\Bigg] \\
&+ 2\mathbb{E}\Bigg[\int_{\Lambda^l}\sum_{\alpha\in M_t}\sum_{\rho\in M_t^c}(a_\alpha(P_i) - a_\alpha(\widehat{P}_i))a_\rho(P_i) \\
&\qquad \varphi_\alpha(x)\varphi_\rho(x)\mathrm{d}x\Bigg]
\end{aligned}
$$

$$+\mathbb{E}\left[\int_{\Lambda^l}\sum_{\alpha\in M_t^c}\sum_{\rho\in M_t^c} a_\alpha(P_i)a_\rho(P_i)\varphi_\alpha(x)\varphi_\rho(x)\mathrm{d}x\right]$$
$$=\mathbb{E}\left[\sum_{\alpha\in M_t}(a_\alpha(P_i)-a_\alpha(\widehat{P}_i))^2\right]+\mathbb{E}\left[\sum_{\alpha\in M_t^c}a_\alpha^2(P_i)\right], \quad (27)$$

where the last line follows from the orthonormality of $\{\varphi\}_{\alpha\in\mathbb{Z}}$. Furthermore, note that $\forall P_i\in\mathcal{I}$:

$$\begin{aligned}\sum_{\alpha\in M_t^c}a_\alpha^2(P_i) =&\frac{1}{t^2}\sum_{\alpha\in M_t^c}t^2a_\alpha^2(P_i)\\ \leq&\frac{1}{t^2}\sum_{\alpha\in\mathbb{Z}}\kappa_\alpha^2(\nu,\gamma)a_\alpha^2(P_i)\\ \leq&\frac{A}{t^2}.\end{aligned} \quad (28)$$

Also,

$$\mathbb{E}\left[(a_\alpha(P_i)-a_\alpha(\widehat{P}_i))^2\right]=\left(\mathbb{E}\left[a_\alpha(\widehat{P}_i)\right]-a_\alpha(P_i)\right)^2 + \mathrm{Var}\left[a_\alpha(\widehat{P}_i)\right].$$

Clearly, $a_\alpha(\widehat{P}_i)$ is unbiased from (9). Also,

$$\begin{aligned}\mathrm{Var}\left[a_\alpha(\widehat{P}_i)\right] =&\frac{1}{n_i^2}\sum_{j=1}^{n_i}\mathrm{Var}\left[\varphi_\alpha(X_{ij})\right]\\ \leq&\frac{n_i\varphi_{\max}^2}{n_i^2}\\ =&O(n_i^{-1}),\end{aligned}$$

where $\varphi_{\max}\equiv\max_{\alpha\in\mathbb{Z}^l}\|\varphi_\alpha\|_\infty$. Thus,

$$\mathbb{E}\left[\|p_i-\tilde{p}_i\|_2^2\right]\leq\frac{C_1|M_t|}{n_i}+\frac{C_2}{t^2}.$$

First note that if we have a bound $\forall\alpha\in M_t$, $|\alpha_i|\leq c_i$ then $|M_t|\leq\prod_{i=1}^l(2c_i+1)$, by a simple counting argument. Let $\lambda=\mathrm{argmin}_i\nu_i^{2\gamma_i}$. For $\alpha\in M_t$ we have:

$$\sum_{i=1}^l|\alpha_i|^{2\gamma_i}\leq\frac{1}{\nu_\lambda^{2\gamma_\lambda}}\sum_{i=1}^l(\nu_i|\alpha_i|)^{2\gamma_i}=\frac{\kappa_\alpha^2(\nu,\gamma)}{\nu_\lambda^{2\gamma_\lambda}}\leq\frac{t^2}{\nu_\lambda^{2\gamma_\lambda}},$$

and

$$|\alpha_i|^{2\gamma_i}\leq\sum_{i=1}^l|\alpha_i|^{2\gamma_i}\leq t^2\nu_\lambda^{-2\gamma_\lambda}\implies|\alpha_i|\leq\nu_\lambda^{-\frac{\gamma_\lambda}{\gamma_i}}t^{\frac{1}{\gamma_i}}.$$

Thus, $|M_t|\leq\prod_{i=1}^l(2\nu_\lambda^{-\frac{\gamma_\lambda}{\gamma_i}}t^{\frac{1}{\gamma_i}}+1)$. Thus, $|M_t|=O\left(t^{\gamma^{-1}}\right)$ where $\gamma^{-1}=\sum_{j=1}^l\gamma_j^{-1}$. Hence,

$$\frac{\partial}{\partial t}\left[\frac{C_1t^{\gamma^{-1}}}{n_i}+\frac{C_2}{t^2}\right]=\frac{C_1't^{\gamma^{-1}-1}}{n_i}-C_2't^{-3}=0\implies$$
$$t=Cn^{\frac{1}{2+\gamma^{-1}}}\implies$$
$$\mathbb{E}\left[\|p_i-\tilde{p}_i\|_2^2\right]\leq\frac{C_1|M_t|}{n_i}+\frac{C_2}{t^2}=O\left(n_i^{-\frac{2}{2+\gamma^{-1}}}\right).$$

Furthermore, by (27) we may see that for $G_i\in\mathcal{I}$, if

$$\bar{g}_i=\sum_{\alpha\in\mathbb{Z}}a_\alpha(G_i)\varphi_\alpha,$$

then

$$\mathbb{E}\left[\|g_i-\bar{g}_i\|_2^2\right]=O\left(n_i^{-\frac{2}{2+\gamma^{-1}}}\right).$$