*Submitted to the Annals of Statistics*
arXiv: arXiv:0000.0000

# SUPPLEMENTARY MATERIAL TO "CONCENTRATION OF TEMPERED POSTERIORS AND OF THEIR VARIATIONAL APPROXIMATIONS"

BY PIERRE ALQUIER* AND JAMES RIDGWAY†

*RIKEN AIP* and Capital Fund Management†*

In this document, we provide the toy example announced in the paper, where the MLE (and thus the MAP) are not defined. Then, we show that there is a variational approximation that leads to a consistent estimator. This also implies that the tempered posterior is also consistent in this case.

## Consistency of variational approximations in a model where the MAP and the MLE are not defined.

*Description of the model.* We define the statistical model

$$P_\theta = \frac{1}{2}\mathcal{N}(m,\sigma^2) + \frac{1}{2}\mathcal{N}(0,1)$$

for $\theta = (m,\sigma^2) \in \Theta = \mathbb{R}\times(0,1)$. Let $g_\theta$ denote the density of $P_\theta$ with respect to the Lebesgue measure.

The prior $\pi$ is given by: $m \sim \mathcal{N}(0,1)$ and $\sigma^2 \sim \mathcal{U}(0,1)$ (uniform distribution).

*Non-existence of the MLE.* It is easy to check that when $X_1,\dots,X_n$ are i.i.d from $P_{(m_0,\sigma_0^2)}$ then the likelihood function

$$\begin{aligned} L_n(m,\sigma^2) &= \prod_{i=1}^n g_{(m,\sigma^2)}(X_i) \\ &= \frac{1}{(2\sqrt{2\pi})^n}\prod_{i=1}^n \left[\frac{\exp\left(-\frac{(X_i-m)^2}{2\sigma^2}\right)}{\sqrt{\sigma^2}} + \exp\left(-\frac{X_i^2}{2}\right)\right] \\ &\geq \frac{1}{(2\sqrt{2\pi})^n}\frac{\exp\left(-\frac{(X_i-m)^2}{2\sigma^2}\right)}{\sqrt{\sigma^2}}\exp\left(-\sum_{j\neq i}\frac{X_j^2}{2}\right) \end{aligned}$$

satisfies, for any $i$,

$$L_n(X_i,\sigma^2) \xrightarrow[\sigma^2\to 0]{} \infty.$$

Thus, the MLE is not defined. For the same reason, the MAP does not exist either.

*A variational approximation family.* We define a family $\mathcal{F}$ that will lead to a consistent estimation. Note that in this case, the variational approximation is not meant to be helpful for computational purposes. On the other hand, it is a very natural one. Indeed, the problem with the MLE is that the likelihood has huge variations around each $X_i$. We will here smooth these variations, creating a kind of maximum of the local mean value of the likelihood. Take the set $\mathcal{F}$ as the set of uniform distributions $\rho_{a,b,c,d}$ over $[a-c,a+c]\times[b-d,b]$, with $(a,b,c,d)\in\mathcal{P}=\{(a,b,c,d)\in\mathbb{R}^4 : b\in(0,1), c>0, 0<d<b\}$. So we remind that the estimator is then defined as $\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)=\rho_{\hat{a},\hat{b},\hat{c},\hat{d}}$ where

$$(\hat{a},\hat{b},\hat{c},\hat{d}) = \operatorname*{arg\,min}_{(a,b,c,d)\in\mathcal{P}} \left\{ -\alpha \int \sum_{i=1}^{n} \log g_{(m,\sigma^2)}(X_i)\rho_{a,b,c,d}(\mathrm{d}(m,\sigma^2)) + \mathcal{K}(\rho_{a,b,c,d},\pi) \right\}.$$

Note that the family $\mathcal{F}$ is inspired by Catoni's point of view [2] to use a "perturbed MLE" in PAC-Bayesian bounds. An application of Theorem 2.6 leads to the following result.

Proposition 0.1. *For any $\alpha\in(0,1)$,*

$$\mathbb{E}\left[\int D_\alpha(P_\theta,P_{\theta_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \leq \frac{1.5\log\left(\frac{2n}{\sigma_0^2}\right)+m_0^2+1.23}{n(1-\alpha)}.$$

As a corollary, we also have that the tempered posterior $\pi_{n,\alpha}(\cdot|X_1^n)$ satisfies the same inequality.

*Proof of Proposition 0.1.* Theorem 2.6 gives

$$\mathbb{E}\left[\int D_\alpha(P_\theta,P_{\theta_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \leq \inf_{(a,b,c,d)\in\mathcal{P}} \left\{ \frac{\alpha}{1-\alpha}\int \mathcal{K}(P_{(m_0,\sigma_0^2)},P_{(m,\sigma)})\rho_{a,b,c,d}(\mathrm{d}(m,\sigma^2)) + \frac{\mathcal{K}(\rho_{a,b,c,d},\pi)}{n(1-\alpha)} \right\}. \tag{0.1}$$

We then have:

$$\mathcal{K}(P_{(m_0,\sigma_0^2)},P_{(m,\sigma^2)}) = \mathcal{K}\left(\frac{1}{2}\mathcal{N}(m_0,\sigma_0^2)+\frac{1}{2}\mathcal{N}(0,1), \frac{1}{2}\mathcal{N}(m,\sigma^2)+\frac{1}{2}\mathcal{N}(0,1)\right)$$

$$\leq \frac{1}{2}\mathcal{K}\left(\mathcal{N}(m_0,\sigma_0^2),\mathcal{N}(m,\sigma^2)\right) \text{ (Theorem 11 in [5])}$$
$$= \frac{1}{4}\left[\frac{(m_0-m)^2}{\sigma^2} + \log\left(\frac{\sigma^2}{\sigma_0^2}\right) + \frac{\sigma_0^2-\sigma^2}{\sigma^2}\right].$$

Assume that $m \in [m_0 - \sqrt{\delta\sigma_0^2}, m_0 + \sqrt{\delta\sigma_0^2}]$ and $\sigma^2 \in [\sigma_0^2 - \delta\sigma_0^2, \sigma_0^2]$ for some $0 < \delta < 1$. Then:

$$\mathcal{K}(P_{(m_0,\sigma_0^2)}, P_{(m,\sigma^2)}) \leq \frac{1}{4}\left[\frac{\delta\sigma_0^2}{\sigma^2} + \frac{\delta\sigma_0^2}{\sigma^2}\right] \leq \frac{\delta\sigma_0^2}{2(\sigma_0^2 - \delta\sigma_0^2)} = \frac{\delta}{2(1-\delta)}.$$

This implies that for any $\delta \in (0,1)$,

$$\int \mathcal{K}(P_{(m_0,\sigma_0^2)}, P_{(m,\sigma^2)})\rho_{m_0,\sigma_0^2,\sqrt{\delta\sigma_0^2},\delta\sigma_0^2}(\mathrm{d}(m,\sigma^2)) \leq \frac{\delta}{2(1-\delta)}.$$

On the other hand,

$$\begin{aligned}
\mathcal{K}(\rho_{m_0,\sigma_0^2,\sqrt{\delta\sigma_0^2},\delta\sigma_0^2}, \pi) &= \mathcal{K}(\mathcal{U}(m_0 - \sqrt{\delta\sigma^2}, m_0 + \sqrt{\delta\sigma^2}), \mathcal{N}(0,1)) \\
&\quad + \mathcal{K}(\mathcal{U}(\sigma_0^2 - \delta\sigma_0^2, \sigma_0^2), \mathcal{U}(0,1)) \\
&= \frac{1}{2\sqrt{\delta\sigma_0^2}} \int_{m_0-\sqrt{\delta\sigma_0^2}}^{m_0+\sqrt{\delta\sigma_0^2}} \log\left(\frac{\sqrt{2\pi}}{2\sqrt{\delta\sigma_0^2}\exp\left(\frac{-x^2}{2}\right)}\right)\mathrm{d}x \\
&\quad + \log\left(\frac{1}{\delta\sigma_0^2}\right) \\
&\leq \frac{1}{2\sqrt{\delta\sigma_0^2}} \int_{-\sqrt{\delta\sigma_0^2}}^{\sqrt{\delta\sigma_0^2}} \left[\frac{(m_0+x)^2}{2} + \log\left(\sqrt{\frac{\pi}{2\delta\sigma_0^2}}\right)\right]\mathrm{d}x \\
&\quad + \log\left(\frac{1}{\delta\sigma_0^2}\right) \\
&\leq m_0^2 + \delta\sigma_0^2 + \frac{1}{2}\log\left(\frac{\pi}{2}\right) + \frac{3}{2}\log\left(\frac{1}{\delta\sigma_0^2}\right).
\end{aligned}$$

Plugging everything into (0.1) gives:

$$\begin{aligned}
&\mathbb{E}\left[\int D_\alpha(P_\theta, P_{\theta_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \\
&\qquad \leq \inf_{\delta>0}\left[\frac{\alpha\delta}{2(1-\alpha)(1-\delta)} + \frac{m_0^2 + \delta\sigma_0^2 + \frac{1}{2}\log\left(\frac{\pi}{2}\right) + \frac{3}{2}\log\left(\frac{1}{\delta\sigma_0^2}\right)}{n(1-\alpha)}\right].
\end{aligned}$$

The value $\delta = 1/(2n)$ gives, using $(1-\delta) > 1/2$ to simplify things,

$$\mathbb{E}\left[\int D_\alpha(P_\theta, P_{\theta_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \leq \frac{\alpha}{2n(1-\alpha)} + \frac{m_0^2 + \frac{\sigma_0^2}{2n} + \frac{1}{2}\log\left(\frac{\pi}{2}\right) + \frac{3}{2}\log\left(\frac{2n}{\sigma_0^2}\right)}{n(1-\alpha)} \leq \frac{0.5}{n(1-\alpha)} + \frac{m_0^2 + 0.5 + 0.23 + 1.5\log\left(\frac{2n}{\sigma_0^2}\right)}{n(1-\alpha)}.$$

**Proof of Theorem 4.1.** Fix $B > 0$, $r \geq 1$ and any pair $(\bar{U}, \bar{V}) \in \mathcal{M}_{r,B}$ and define for $\delta \in (0, B)$ that will be chosen later,

$$\rho_n(\mathrm{d}U, \mathrm{d}V, \mathrm{d}\gamma) \propto \mathbf{1}(\|U - \bar{U}\|_\infty \leq \delta, \|U - \bar{U}\|_\infty \leq \delta)\pi(\mathrm{d}U, \mathrm{d}V, \mathrm{d}\gamma).$$

Note that it can be factorized so it belongs to the family $\mathcal{F}$.

We adapt the calculations from [1, 3] to our context. First, note that

$$\mathcal{K}(P_M, P_{UV^t}) = \frac{1}{mp}\sum_{i=1}^{m}\sum_{j=1}^{p}\frac{(M_{i,j} - (UV^t)_{i,j})^2}{2\sigma^2} = \frac{\|M - UV^t\|_F^2}{2\sigma^2 mp}$$

and that for any $(U, V)$ in the support of $\rho_n$ we have

$$\begin{aligned}
\|M - UV^t\|_F &= \|M - \bar{U}\bar{V}^t + \bar{U}\bar{V}^t - \bar{U}V^t + \bar{U}V^t - UV^t\|_F \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + \|\bar{U}\bar{V}^t - \bar{U}V^t\|_F + \|\bar{U}V^t - UV^t\|_F \\
&= \|M - \bar{U}\bar{V}^t\|_F + \|\bar{U}(\bar{V}^t - V^t)\|_F + \|(\bar{U} - U)V^t\|_F \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + \|\bar{U}\|_F\|\bar{V} - V\|_F + \|\bar{U} - U\|_F\|V^t\|_F \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + mp\|\bar{U}\|_\infty^{1/2}\|\bar{V} - V\|_\infty^{1/2} + mp\|V\|_\infty^{1/2}\|\bar{U} - U\|_\infty^{1/2} \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + mp(B^{1/2}\delta^{1/2} + (B+\delta)^{1/2}\delta^{1/2}) \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + 2mp\delta^{1/2}(B+\delta)^{1/2} \\
&\leq \|M - \bar{U}\bar{V}^t\|_F + 2^{3/2}mp\delta^{1/2}B^{1/2} = \|M - \bar{U}\bar{V}^t\|_F + B/n
\end{aligned}$$

with the choice $\delta = B/[8(nmp)^2]$ which satisfies $0 < \delta < B$. Then, we derive

$$\begin{aligned}
\mathcal{K}(\rho_n, \pi) &= \log\frac{1}{\pi\left(\|U - \bar{U}\|_\infty \leq \delta, \|U - \bar{U}\|_\infty \leq \delta\right)} \\
&= \log\frac{1}{\int \pi\left(\|U - \bar{U}\|_\infty \leq \delta, \|U - \bar{U}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma)} \\
&= \log\frac{1}{\int \pi\left(\|U - \bar{U}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma)} + \log\frac{1}{\int \pi\left(\|V - \bar{V}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma)}
\end{aligned}$$

$$= \log \frac{1}{\int_E \pi\left(\|U - \bar{U}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma)} + \log \frac{1}{\int_E \pi\left(\|V - \bar{V}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma)}$$

for any event $E$. We actually take $E = \{\gamma_1, \ldots, \gamma_r \in [B^2, 2B^2], \gamma_{r+1}, \ldots, \gamma_K \in [s, 2s]\}$ and $s \in (0, B^2)$ is to be chosen later. Then note that

$$\begin{aligned}
&\pi\left(\|U - \bar{U}\|_\infty \leq \delta|\gamma\right) = \pi\left(\forall i, k : |U_{i,k} - \bar{U}_{i,k}| \leq \delta|\gamma\right) \\
&= \prod_{i=1}^{m}\prod_{k=1}^{r} \pi\left(|U_{i,k} - \bar{U}_{i,k}| \leq \delta|\gamma_k\right)\pi\left(\max_{i=1}^{m}\max_{k=r+1}^{K}|U_{i,k}| \leq \delta\middle|\gamma_{r+1}, \ldots, \gamma_K\right).
\end{aligned}$$

First,

$$\begin{aligned}
&\pi\left(\max_{i=1}^{m}\max_{k=r+1}^{K}|U_{i,k}| \leq \delta\middle|\gamma_{r+1}, \ldots, \gamma_K\right) \\
&= 1 - \pi\left(\max_{i=1}^{m}\max_{k=r+1}^{K}|U_{i,k}| > \delta\middle|\gamma_{r+1}, \ldots, \gamma_K\right) \\
&\geq 1 - \pi\left(\sum_{i=1}^{m}\sum_{k=r+1}^{K}|U_{i,k}| > \delta\middle|\gamma_{r+1}, \ldots, \gamma_K\right) \\
&\geq 1 - \frac{\sum_{i=1}^{m}\sum_{k=r+1}^{K}\pi\left(|U_{i,k}|\,|\,\gamma_k\right)}{\delta} \\
&\geq 1 - \frac{m(K-r)\max_{k\geq r+1}\sqrt{\gamma_k}}{\delta} \\
&\geq 1 - \frac{mK\sqrt{2s}}{\delta} = 1/2
\end{aligned}$$

with $s = \frac{1}{2}\left(\frac{\delta}{2(m\vee p)K}\right)^2$ which satisfies $0 < s < B^2$. Then, for $k \leq r$,

$$\begin{aligned}
\pi\left(|U_{i,k} - \bar{U}_{i,k}| \leq \delta|\gamma_k\right) &= \frac{1}{\sqrt{2\pi\gamma_k}}\int_{\bar{U}_{i,k}-\delta}^{\bar{U}_{i,k}+\delta}\exp\left(-\frac{x^2}{2\gamma_k}\right)\mathrm{d}x \\
&\geq \frac{2\delta\exp\left(-\frac{(B+\delta)^2}{2\gamma_k}\right)}{\sqrt{2\pi\gamma_k}} \\
&\geq \frac{\delta\exp\left(-\frac{(B+\delta)^2}{2B^2}\right)}{B\sqrt{\pi}} \text{ as } B^2 \leq \gamma_k \leq 2B^2 \\
&\geq \frac{\delta\exp(-2)}{B\sqrt{\pi}} \text{ as } \delta < 1 \leq B
\end{aligned}$$

and so

$$\prod_{i=1}^{m}\prod_{k=1}^{r}\pi\left(|U_{i,k} - \bar{U}_{i,k}| \leq \delta|\gamma_k\right) \geq \left(\frac{\delta}{B\sqrt{\pi}}\right)^{mr}\exp(-2mr).$$

Finally

$$\int_E \pi\left(\|U-\bar{U}\|_\infty \leq \delta|\gamma\right)\pi(\mathrm{d}\gamma) \geq \int_E \frac{1}{2}\left(\frac{\delta}{B\sqrt{\pi}}\right)^{mr} \exp\left(-2mr\right)\pi(\mathrm{d}\gamma)$$
$$= \frac{1}{2}\left(\frac{\delta}{B\sqrt{\pi}}\right)^{mr} \exp\left(-2mr\right)\pi(\gamma\in E)$$

and it remains to lower bound

$$\begin{aligned}
\pi(\gamma\in E) &= \left(\prod_{k=1}^{r}\pi(1\leq\gamma_k\leq 2)\right)\left(\prod_{k=r+1}^{K}\pi(s\leq\gamma_k\leq 2s)\right) \\
&= \left(\frac{b^a}{\Gamma(a)}\right)^K \left[\int_{B^2}^{2B^2} x^{-a-1}\exp\left(-\frac{b}{x}\right)\mathrm{d}x\right]^r \left[\int_s^{2s} x^{-a-1}\exp\left(-\frac{b}{x}\right)\mathrm{d}x\right]^{K-r} \\
&\geq \left(\frac{b^a}{\Gamma(a)}\right)^K \left[B^2(2B^2)^{-a-1}\exp\left(-\frac{b}{B^2}\right)\right]^r \left[s(2s)^{-a-1}\exp\left(-\frac{b}{s}\right)\right]^{K-r} \\
&= \left(\frac{b^a}{2^{a+1}\Gamma(a)}\right)^K \exp\left[-\frac{b}{B^2}r-\frac{b}{s}(K-r)\right](B^2)^{-(a+1)r}s^{-a(K-r)} \\
&\geq \left(\frac{b^a}{(B^2)^a 2^{a+1}\Gamma(a)}\right)^K \exp\left[-\frac{Kb}{s}\right]
\end{aligned}$$

as $s<B^2$. So, finally,

$$\mathcal{K}(\rho_n,\pi) \leq r(m+p)\log\left(\frac{B\sqrt{\pi}\exp(2)}{\delta}\right)+K\left[\log\left(\frac{2^{a+1}\Gamma(a)(B^2)^a}{b^a}\right)+\frac{b}{s}\right]+2\log(2).$$

The choice $b=s$ leads to

$$\begin{aligned}
\mathcal{K}(\rho_n,\pi) &\leq r(m+p)\log\left(\frac{B\sqrt{\pi}\exp(2)}{\delta}\right)+K\log\left(\frac{\mathrm{e}2^{a+1}\Gamma(a)(B^2)^a}{s^a}\right)+2\log(2) \\
&\leq r(m+p)\log\left(8\sqrt{\pi}\exp(2)(nmp)^2\right)+4aK\log\left(nmp\right) \\
&\quad + K\log(\mathrm{e}2^{10a+1}\Gamma(a))+2\log(2)
\end{aligned}$$

where we replaced $\delta$ and $s$ by their respective value. In order to keep the expressions as simple as possible we can use $K\leq m\vee p\leq m+p\leq r(m+p)$ and $2\leq m+p\leq r(m+p)$ to get

$$\mathcal{K}(\rho_n,\pi)\leq 2(1+2a)r(m+p)\left[\log(nmp)+\underbrace{\log(8\sqrt{\pi}\Gamma(a)2^{10a+1})+3}_{=:\mathcal{C}(a)}\right].$$

We are now in position to apply Theorem 2.6. Then

$$\mathbb{E}\left[\int D_{\alpha}(P_{\theta},P_{\theta_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \leq \frac{\alpha}{1-\alpha}\int \mathcal{K}(P_{\theta_0},P_{\theta})\rho_n(\mathrm{d}\theta)+\frac{\mathcal{K}(\rho_n,\pi)}{n(1-\alpha)}$$
$$\leq \frac{\alpha}{1-\alpha}\frac{\left[\|M-\bar{U}\bar{V}^t\|_F+\frac{\sqrt{B}}{n}\right]^2}{2\sigma^2 mp} + \frac{2(1+\alpha)(1+2a)r(m+p)\left[\log(nmp)+\mathcal{C}(a,B)\right]}{n(1-\alpha)}.$$

**Proof of Corollary 4.2.** We start from (4.1). Under the boundedness assumption on $M_0$ it is obvious that $\forall M$, $d_{\alpha,\sigma}(M,M_0)\geq d_{\alpha,\sigma}(\mathrm{clip}_C(M),M_0)$, so

$$\mathbb{E}\left[\int d_{\alpha,\sigma}(\mathrm{clip}_C(M),M_0)\tilde{\pi}_{n,\alpha}(\mathrm{d}M|X_1^n)\right] \leq \frac{2(1+\alpha)(1+2a)r(m+p)\left[\log(nmp)+\mathcal{C}(a)+\frac{\alpha B}{2\sigma^2 mp}\right]}{n(1-\alpha)}. \tag{0.2}$$

Fix $M$ and for short, put $N=\mathrm{clip}_C(M)$. We have:

$$d_{\alpha,\sigma}(N,M_0)=\frac{-1}{1-\alpha}\log\left[\frac{1}{mp}\sum_{i=1}^{m}\sum_{j=1}^{p}\exp\left(\frac{\alpha(\alpha-1)(N_{i,j}-(M_0)_{i,j})^2}{2\sigma^2}\right)\right]$$
$$\geq \frac{1}{1-\alpha}\left[1-\frac{1}{mp}\sum_{i=1}^{m}\sum_{j=1}^{p}\exp\left(\frac{\alpha(\alpha-1)(N_{i,j}-(M_0)_{i,j})^2}{2\sigma^2}\right)\right].$$

By assumption, $(N_{i,j}-(M_0)_{i,j})^2/(2\sigma^2)\leq(2C)^2/(2\sigma^2)=2C^2/\sigma^2$. Straightforward derivations show that for any $x\in[0,2C^2/\sigma^2]$ we have

$$1-\left(\frac{\sigma^2[1-\exp(2C^2\alpha(\alpha-1)/\sigma^2)]}{2C^2}\right)x\geq\exp(\alpha(\alpha-1)x),$$

this leads to

$$d_{\alpha,\sigma}(N,M_0)\geq\frac{1}{1-\alpha}\frac{\sigma^2[1-\exp(2C^2\alpha(\alpha-1)/\sigma^2)]}{2C^2}\frac{\|N-M_0\|_F^2}{2\sigma^2}.$$

Pluging this into (0.2) gives the result claimed.

**Proof of Theorem 5.1.** Let $(\beta_k^0)$ denote the coefficients of $f^0$: $f_0 = \sum_{k=1}^{\infty} \beta_k^0 \varphi_k$. Theorem 2.6 gives:

$$
\begin{aligned}
&\mathbb{E}\left[\int D_\alpha(P_f, P_{f_0}) \tilde{\pi}_{n,\alpha}(\mathrm{d}\theta | X_1^n)\right] \\
&\leq \frac{1}{1-\alpha} \inf_{K\geq 1} \inf_{\mathbf{m}, s^2} \left\{ \frac{\alpha}{2} \int\int_{-1}^{1} \left( f_0 - \sum_{k=1}^{K} \beta_k \varphi_k \right)^2 \Phi(\mathrm{d}\beta, \mathbf{m}, s^2) \right. \\
&\qquad \left. + \frac{\sum_{k=1}^{K} \frac{1}{2}\left[\log\left(\frac{1}{s^2}\right) + s^2 + m_k^2 - 1\right] + K\log(2)}{n} \right\} \\
&= \frac{1}{1-\alpha} \inf_{K\geq 1} \inf_{\mathbf{m}, s^2} \left\{ \frac{\alpha}{2} \int\int_{-1}^{1} \left( \sum_{k=1}^{\infty} \beta_k^0 - \sum_{k=1}^{K} \beta_k \varphi_k \right)^2 \Phi(\mathrm{d}\beta, \mathbf{m}, s^2) \right. \\
&\qquad \left. + \frac{\sum_{k=1}^{K} \frac{1}{2}\left[\log\left(\frac{1}{s^2}\right) + s^2 + m_k^2 - 1\right] + K\log(2)}{n} \right\} \\
&= \frac{1}{1-\alpha} \inf_{K\geq 1} \inf_{\mathbf{m}, s^2} \left\{ \frac{\alpha}{2} \sum_{k=1}^{K} (m_k - \beta_k^0)^2 + \frac{\alpha K s^2}{2} + \frac{\alpha}{2} \sum_{k=K+1}^{\infty} (\beta_k^0)^2 \right. \\
&\qquad \left. + \frac{\sum_{k=1}^{K} \frac{1}{2}\left[\log\left(\frac{1}{s^2}\right) + s^2 + m_k^2 - 1\right] + K\log(2)}{n} \right\}.
\end{aligned}
$$

The choice $(m_1, \dots, m_K) = (\beta_1, \dots, \beta_K)$ gives:

$$
\begin{aligned}
\mathbb{E}\left[\int D_\alpha(P_f, P_{f_0}) \tilde{\pi}_{n,\alpha}(\mathrm{d}\theta | X_1^n)\right] &\leq \frac{1}{1-\alpha} \inf_{K\geq 1} \left\{ \frac{\alpha K s^2}{2} + \frac{\alpha}{2} \sum_{k=K+1}^{\infty} (\beta_k^0)^2 \right. \\
&\qquad \left. + \frac{\sum_{k=1}^{K} \frac{1}{2}\left[\log\left(\frac{1}{s^2}\right) + s^2 + (\beta_k^0)^2 - 1\right] + K\log(2)}{n} \right\}.
\end{aligned}
$$

From Chapter 1 in [4] we know that $f_0 \in \mathcal{W}(r, C^2)$ implies $\sum_{k=K+1}^{\infty} (\beta_k^0)^2 \leq \Lambda(r, C) K^{-2r}$ for some $\Lambda(k, C)$. Moreover, $\sum_{k=1}^{K} (\beta_k^0)^2 \leq \sum_{k=1}^{\infty} k^{2r} (\beta_k^0)^2 \leq C^2$. So finally:

$$
\begin{aligned}
\mathbb{E}\left[\int D_\alpha(P_f, P_{f_0}) \tilde{\pi}_{n,\alpha}(\mathrm{d}\theta | X_1^n)\right] &\leq \frac{1}{1-\alpha} \inf_{K\geq 1} \left\{ \frac{\alpha \Lambda(r, C)}{2} K^{-2r} \right. \\
&\qquad \left. + \frac{\alpha K s^2}{2} + \frac{K\left[\log(2) + \frac{s^2}{2} + \frac{\log\left(\frac{1}{s^2}\right)}{2}\right] + C^2}{n} \right\}.
\end{aligned}
$$

The choice $s^2 = 1/n$ leads to

$$\mathbb{E}\left[\int D_\alpha(P_f, P_{f_0})\tilde{\pi}_{n,\alpha}(\mathrm{d}\theta|X_1^n)\right] \leq \frac{1}{1-\alpha}\inf_{K\geq 1}\left\{\frac{\alpha\Lambda(r,C)}{2}K^{-2r} + \frac{K\left[\log(2)+\frac{\alpha}{2}+\frac{1}{2n}+\frac{\log(n)}{2}\right]+C^2}{n}\right\}.$$

The choice $K = \lceil (n/\log(n))^{1/(2r+1)})\rceil$ leads to the result.

**References.**


[1] P. Alquier and B. Guedj. An oracle inequality for quasi-Bayesian non-negative matrix factorization. *Mathematical Methods of Statistics*, 26(1):55–67, 2017.

[2] O. Catoni. *Statistical Learning Theory and Stochastic Optimization.* Saint-Flour Summer School on Probability Theory 2001 (Jean Picard ed.), Lecture Notes in Mathematics. Springer, 2004.

[3] V. Cottet and P. Alquier. 1-Bit matrix completion: PAC-Bayesian analysis of a variational approximation. *Machine Learning*, 107(3), 579–603, 2018.

[4] A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer Series in Statistics, 2009.

[5] T. Van Erven and P. Harremos. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.



Pierre Alquier
RIKEN AIP
Nihonbashi 1-chome Mitsui Building, 15th floor,
1-4-1 Nihonbashi, Chuo-ku, Tokyo,
103-0027, JAPAN
E-mail: pierre.alquier.stat@gmail.com

James Ridgway
Capital Fund Management
23 rue de l'Université
75007 Paris
FRANCE
E-mail: james.lp.ridgway@gmail.com