Chapter 2

Comparing posterior distributions to Gibbs priors

2.1. Bounds relative to a Gibbs distribution

We now come to an approach to relative bounds whose performance can be analysed with PAC-Bayesian tools.

The empirical bounds at the end of the previous chapter involve taking suprema in $\theta \in \Theta$, and replacing the *expected margin function* φ with some empirical counterparts $\overline{\varphi}$ or $\widetilde{\varphi}$, which may prove unsafe when using very complex classification models.

We are now going to focus on the control of the divergence $\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr]$. It is already obvious, we hope, that controlling this divergence is the crux of the matter, and that it is a way to upper bound the mutual information between the training sample and the parameter, which can be expressed as $\mathcal{K}\bigl[\rho, \mathbb{P}(\rho)\bigr] = \mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] - \mathcal{K}\bigl[\mathbb{P}(\rho), \pi_{\exp(-\beta R)}\bigr]$, as explained on page 14.

Through the identity

$$(2.1) \quad \mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] = \beta\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] + \mathcal{K}(\rho, \pi) - \mathcal{K}\bigl[\pi_{\exp(-\beta R)}, \pi\bigr],$$

we see that the control of this divergence is related to the control of the difference $\rho(R) - \pi_{\exp(-\beta R)}(R)$. This is the route we will follow first.

Thus comparing any posterior distribution with a Gibbs prior distribution will provide a first way to build an estimator which can be proved to reach adaptively the best possible asymptotic error rate under Mammen and Tsybakov margin assumptions and parametric complexity assumptions (at least as long as orders of magnitude are concerned, we will not discuss the question of asymptotically optimal constants).

Then we will provide an empirical bound for the Kullback divergence $\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr]$ itself. This will serve to address the question of model selection, which will be achieved by comparing the performance of two posterior distributions possibly supported by two different models. This will also provide a second way to build estimators which can be proved to be adaptive under Mammen and Tsybakov margin assumptions and parametric complexity assumptions (somewhat weaker than with the first method).

Finally, we will present two-step localization strategies, in which the performance of the posterior distribution to be analysed is compared with a *two-step* Gibbs prior.

2.1.1. Comparing a posterior distribution with a Gibbs prior. Similarly to Theorem 1.4.3 (page 37) we can prove that for any prior distribution $\widetilde{\pi} \in \mathcal{M}_+^1(\Theta)$,

$$\mathbb{P}\biggl\{\widetilde{\pi}\otimes\widetilde{\pi}\biggl\{\exp\Bigl[-N\log(1-N\tanh(\tfrac{\gamma}{N})R') \\ -\gamma r' - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\Bigr]\biggr\}\biggr\} \leq 1. \tag{2.2}$$

Replacing $\widetilde{\pi}$ with $\pi_{\exp(-\beta R)}$ and considering the posterior distribution $\rho\otimes\pi_{\exp(-\beta R)}$, provides a starting point in the comparison of ρ with $\pi_{\exp(-\beta R)}$; we can indeed state with $\mathbb{P}$ probability at least $1-\epsilon$ that

$$-N\log\Bigl\{1-\tanh(\tfrac{\gamma}{N})\Bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\Bigr]\Bigr\} \\ \leq \gamma\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr] + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\bigl[\rho\otimes\pi_{\exp(-\beta R)}\bigr](m') \\ + \mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr] - \log(\epsilon). \tag{2.3}$$

Using equation (2.1, page 51) to handle the entropy term, we get

$$-N\log\Bigl\{1-\tanh(\tfrac{\gamma}{N})\Bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\Bigr]\Bigr\} - \beta\bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\bigr] \\ \leq \gamma\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr] + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho\otimes\pi_{\exp(-\beta R)}(m') \\ + \mathcal{K}(\rho,\pi) - \mathcal{K}\bigl[\pi_{\exp(-\beta R)},\pi\bigr] - \log(\epsilon). \tag{2.4}$$

We can then decompose in the right-hand side $\gamma\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr]$ into $(\gamma-\lambda)\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr]+\lambda\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr]$ for some parameter λ to be set later on and use the fact that

$$\lambda\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr] + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho\otimes\pi_{\exp(-\beta R)}(m') \\ + \mathcal{K}(\rho,\pi) - \mathcal{K}\bigl[\pi_{\exp(-\beta R)},\pi\bigr] \\ \leq \lambda\rho(r) + \mathcal{K}(\rho,\pi) + \log\Bigl\{\pi\Bigl[\exp\bigl\{-\lambda r + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} \\ = \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda r)}\bigr] + \log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\},$$

to get rid of the appearance of the unobserved Gibbs prior $\pi_{\exp(-\beta R)}$ in most places of the right-hand side of our inequality, leading to

Theorem 2.1.1. *For any real constants β and γ, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \rightarrow \mathcal{M}_+^1(\Theta)$, for any real constant λ,*

$$\bigl[N\tanh(\tfrac{\gamma}{N})-\beta\bigr]\bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\bigr] \\ \leq -N\log\Bigl\{1-\tanh(\tfrac{\gamma}{N})\Bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\Bigr]\Bigr\} \\ - \beta\bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\bigr] \\ \leq (\gamma-\lambda)\bigl[\rho(r)-\pi_{\exp(-\beta R)}(r)\bigr] + \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda r)}\bigr]$$

$$
\begin{aligned}
&\qquad + \log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} - \log(\epsilon)\\
&= \mathcal{K}\bigl[\rho, \pi_{\exp(-\gamma r)}\bigr]\\
&\quad + \log\Bigl\{\pi_{\exp(-\gamma r)}\Bigl[\exp\bigl\{(\gamma-\lambda)r + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\}\\
&\qquad\qquad - (\gamma-\lambda)\pi_{\exp(-\beta R)}(r) - \log(\epsilon).
\end{aligned}
$$

We would like to have a fully empirical upper bound even in the case when $\lambda \neq \gamma$. This can be done by using the theorem twice. We will need a lemma.

Lemma 2.1.2 *For any probability distribution $\pi \in \mathcal{M}_+^1(\Theta)$, for any bounded measurable functions $g, h : \Theta \to \mathbb{R}$,*

$$
\pi_{\exp(-g)}(g) - \pi_{\exp(-h)}(g) \leq \pi_{\exp(-g)}(h) - \pi_{\exp(-h)}(h).
$$

Proof. Let us notice that

$$
\begin{aligned}
0 \leq \mathcal{K}(\pi_{\exp(-g)}, \pi_{\exp(-h)}) &= \pi_{\exp(-g)}(h) + \log\bigl\{\pi\bigl[\exp(-h)\bigr]\bigr\} + \mathcal{K}(\pi_{\exp(-g)}, \pi)\\
&= \pi_{\exp(-g)}(h) - \pi_{\exp(-h)}(h) - \mathcal{K}(\pi_{\exp(-h)}, \pi) + \mathcal{K}(\pi_{\exp(-g)}, \pi)\\
&= \pi_{\exp(-g)}(h) - \pi_{\exp(-h)}(h) - \mathcal{K}(\pi_{\exp(-h)}, \pi) - \pi_{\exp(-g)}(g) - \log\bigl\{\pi\bigl[\exp(-g)\bigr]\bigr\}.
\end{aligned}
$$

Moreover

$$
-\log\bigl\{\pi\bigl[\exp(-g)\bigr]\bigr\} \leq \pi_{\exp(-h)}(g) + \mathcal{K}(\pi_{\exp(-h)}, \pi),
$$

which ends the proof. □

For any positive real constants β and λ, we can then apply Theorem 2.1.1 to $\rho = \pi_{\exp(-\lambda r)}$, and use the inequality

$$
\frac{\lambda}{\beta}\bigl[\pi_{\exp(-\lambda r)}(r) - \pi_{\exp(-\beta R)}(r)\bigr] \leq \pi_{\exp(-\lambda r)}(R) - \pi_{\exp(-\beta R)}(R) \tag{2.5}
$$

provided by the previous lemma. We thus obtain with $\mathbb{P}$ probability at least $1-\epsilon$

$$
\begin{aligned}
&- N\log\Bigl\{1 - \tanh(\tfrac{\gamma}{N})\tfrac{\lambda}{\beta}\Bigl[\pi_{\exp(-\lambda r)}(r) - \pi_{\exp(-\beta R)}(r)\Bigr]\Bigr\}\\
&\qquad\qquad - \gamma\bigl[\pi_{\exp(-\lambda r)}(r) - \pi_{\exp(-\beta R)}(r)\bigr]\\
&\quad \leq \log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda r)}(m')\bigr\}\Bigr]\Bigr\} - \log(\epsilon).
\end{aligned}
$$

Let us introduce the convex function

$$
F_{\gamma,\alpha}(x) = -N\log\bigl[1 - \tanh(\tfrac{\gamma}{N})x\bigr] - \alpha x \geq \bigl[N\tanh(\tfrac{\gamma}{N}) - \alpha\bigr]x.
$$

With $\mathbb{P}$ probability at least $1-\epsilon$,

$$
\begin{aligned}
-\pi_{\exp(-\beta R)}(r) \leq \inf_{\lambda\in\mathbb{R}_+^*}\Biggl\{&-\pi_{\exp(-\lambda r)}(r)\\
&+ \frac{\beta}{\lambda}F^{-1}_{\gamma,\frac{\beta\gamma}{\lambda}}\Biggl[\log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda r)}(m')\bigr\}\Bigr]\Bigr\}\\
&\qquad\qquad - \log(\epsilon)\Biggr]\Biggr\}.
\end{aligned}
$$

Since Theorem 2.1.1 holds uniformly for any posterior distribution ρ, we can apply it again to some arbitrary posterior distribution ρ. We can moreover make the result uniform in β and γ by considering some atomic measure $\nu \in \mathcal{M}_+^1(\mathbb{R})$ on the real line and using a union bound. This leads to

Theorem 2.1.3. *For any atomic probability distribution on the positive real line $\nu \in \mathcal{M}_+^1(\mathbb{R}_+)$, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$, for any positive real constants β and γ,*

$$\bigl[N\tanh(\tfrac{\gamma}{N})-\beta\bigr]\bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\bigr] \leq F_{\gamma,\beta}\bigl[\rho(R)-\pi_{\exp(-\beta R)}(R)\bigr] \leq B(\rho,\beta,\gamma), \text{ where}$$

$$\begin{aligned}
B(\rho,\beta,\gamma) = \inf_{\substack{\lambda_1\in\mathbb{R}_+,\lambda_1\leq\gamma\\ \lambda_2\in\mathbb{R},\lambda_2>\frac{\beta\gamma}{N}\tanh(\frac{\gamma}{N})^{-1}}} \Biggl\{ & \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda_1 r)}\bigr] \\
& + (\gamma-\lambda_1)\bigl[\rho(r)-\pi_{\exp(-\lambda_2 r)}(r)\bigr] \\
& + \log\Bigl\{\pi_{\exp(-\lambda_1 r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} - \log\bigl[\epsilon\nu(\beta)\nu(\gamma)\bigr] \\
& + (\gamma-\lambda_1)\frac{\beta}{\lambda_2}F^{-1}_{\gamma,\frac{\beta\gamma}{\lambda_2}}\Biggl[\log\Bigl\{ \\
& \qquad \pi_{\exp(-\lambda_2 r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda_2 r)}(m')\bigr\}\Bigr]\Bigr\} \\
& \qquad\qquad - \log\bigl[\epsilon\nu(\beta)\nu(\gamma)\bigr]\Biggr]\Biggr\}
\end{aligned}$$

$$\begin{aligned}
\leq \inf_{\substack{\lambda_1\in\mathbb{R}_+,\lambda_1\leq\gamma\\ \lambda_2\in\mathbb{R},\lambda_2>\frac{\beta\gamma}{N}\tanh(\frac{\gamma}{N})^{-1}}} \Biggl\{ & \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda_1 r)}\bigr] \\
& + (\gamma-\lambda_1)\bigl[\rho(r)-\pi_{\exp(-\lambda_2 r)}(r)\bigr] \\
& + \log\Bigl\{\pi_{\exp(-\lambda_1 r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} \\
& + \frac{\beta}{\lambda_2}\frac{(1-\frac{\lambda_1}{\gamma})}{\bigl[\frac{N}{\gamma}\tanh(\frac{\gamma}{N})-\frac{\beta}{\lambda_2}\bigr]}\log\Bigl\{\pi_{\exp(-\lambda_2 r)}\Bigl[\\
& \qquad \exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda_2 r)}(m')\bigr\}\Bigr]\Bigr\} \\
& \qquad - \Bigl\{1+\frac{\beta}{\lambda_2}\frac{(1-\frac{\lambda_1}{\gamma})}{\bigl[\frac{N}{\gamma}\tanh(\frac{\gamma}{N})-\frac{\beta}{\lambda_2}\bigr]}\Bigr\}\log\bigl[\epsilon\nu(\beta)\nu(\gamma)\bigr]\Biggr\},
\end{aligned}$$

where we have written for short $\nu(\beta)$ and $\nu(\gamma)$ instead of $\nu(\{\beta\})$ and $\nu(\{\gamma\})$.

Let us notice that $B(\rho,\beta,\gamma)=+\infty$ when $\nu(\beta)=0$ or $\nu(\gamma)=0$, the uniformity in β and γ of the theorem therefore necessarily bears on a countable number of values of these parameters. We can typically choose distributions for ν such as the one used in Theorem 1.2.8 (page 13): namely we can put for some positive real ratio $\alpha > 1$

$$\nu(\alpha^k) = \frac{1}{(k+1)(k+2)}, \qquad k \in \mathbb{N},$$

or alternatively, since we are interested in values of the parameters less than N, we can prefer

$$\nu(\alpha^k) = \frac{\log(\alpha)}{\log(\alpha N)}, \qquad 0 \leq k < \frac{\log(N)}{\log(\alpha)}.$$

We can also use such a coding distribution on dyadic numbers as the one defined by equation (1.7, page 15).

Following the same route as for Theorem 1.3.15 (page 30), we can also prove the following result about the deviations under any posterior distribution ρ:

THEOREM 2.1.4 *For any $\epsilon \in)0,1($, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$, with ρ probability at least $1-\xi$,*

$$F_{\gamma,\beta}\bigl[R(\widehat{\theta}) - \pi_{\exp(-\beta R)}(R)\bigr] \leq \inf_{\substack{\lambda_1 \in \mathbb{R}_+, \lambda_1 \leq \gamma, \\ \lambda_2 \in \mathbb{R}, \lambda_2 > \frac{\beta\gamma}{N} \tanh(\frac{\gamma}{N})^{-1}}} \Biggl\{ \log \biggl[\frac{d\rho}{d\pi_{\exp(-\lambda_1 r)}}(\widehat{\theta}\,) \biggr]$$
$$+ (\gamma - \lambda_1)\bigl[r(\widehat{\theta}\,) - \pi_{\exp(-\lambda_2 r)}(r)\bigr]$$
$$+ \log\Bigl\{ \pi_{\exp(-\lambda_1 r)} \Bigl[\exp\bigl\{ N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] m'(\cdot, \widehat{\theta}\,)\bigr\}\Bigr]\Bigr\} - \log\bigl[\epsilon\xi\nu(\beta)\nu(\gamma)\bigr]$$
$$+ (\gamma - \lambda_1)\frac{\beta}{\lambda_2} F^{-1}_{\gamma, \frac{\beta\gamma}{\lambda_2}} \biggl[\log\Bigl\{$$
$$\pi_{\exp(-\lambda_2 r)} \Bigl[\exp\bigl\{ N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] \pi_{\exp(-\lambda_2 r)}(m')\bigr\}\Bigr]\Bigr\}$$
$$- \log\bigl[\epsilon\nu(\beta)\nu(\gamma)\bigr]\biggr]\Biggr\}.$$

The only tricky point is to justify that we can still take an infimum in λ_1 without using a union bound. To justify this, we have to notice that the following variant of Theorem 2.1.1 (page 52) holds: with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$, for any real constant λ,

$$\rho\Bigl\{ F_{\gamma,\beta}\bigl[R - \pi_{\exp(-\beta R)}(R)\bigr]\Bigr\} \leq \mathcal{K}\bigl[\rho, \pi_{\exp(-\gamma r)}\bigr]$$
$$+ \rho\biggl[\inf_{\lambda \in \mathbb{R}} \log\Bigl\{ \pi_{\exp(-\gamma r)} \Bigl[\exp\bigl\{ (\gamma - \lambda) r + N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] m'(\cdot, \widehat{\theta}\,)\bigr\}\Bigr]\Bigr\}$$
$$- (\gamma - \lambda)\pi_{\exp(-\beta R)}(r)\biggr] - \log(\epsilon).$$

We leave the details as an exercise.

2.1.2. THE EFFECTIVE TEMPERATURE OF A POSTERIOR DISTRIBUTION. Using the parametric approximation $\pi_{\exp(-\alpha r)}(r) - \inf_\Theta r \simeq \frac{d_e}{\alpha}$, we get as an order of magnitude

$$B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma) \lesssim -(\gamma - \lambda_1) d_e \bigl[\lambda_2^{-1} - \lambda_1^{-1}\bigr]$$
$$+ 2 d_e \log \frac{\lambda_1}{\lambda_1 - N \log\bigl[\cosh(\frac{\gamma}{N})\bigr] x}$$
$$+ 2 \frac{\beta}{\lambda_2} \frac{(1 - \frac{\lambda_1}{\gamma})}{\bigl[\frac{N}{\gamma}\tanh(\frac{\gamma}{N}) - \frac{\beta}{\lambda_2}\bigr]} d_e \log \left(\frac{\lambda_2}{\lambda_2 - N \log\bigl[\cosh(\frac{\gamma}{N})\bigr] x}\right)$$
$$+ 2N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] \left[1 + \frac{\beta}{\lambda_2} \frac{(1 - \frac{\lambda_1}{\gamma})}{\bigl[\frac{N}{\gamma}\tanh(\frac{\gamma}{N}) - \frac{\beta}{\lambda_2}\bigr]}\right] \widetilde{\varphi}(x)$$
$$- \left\{ 1 + \frac{\beta}{\lambda_2} \frac{(1 - \frac{\lambda_1}{\gamma})}{\bigl[\frac{N}{\gamma}\tanh(\frac{\gamma}{N}) - \frac{\beta}{\lambda_2}\bigr]}\right\} \log\bigl[\nu(\beta)\nu(\gamma)\epsilon\bigr].$$

Therefore, if the empirical dimension d_e stays bounded when N increases, we are going to obtain a negative upper bound for any values of the constants $\lambda_1 > \lambda_2 > \beta$, as soon as γ and $\frac{N}{\gamma}$ are chosen to be large enough. This ability to obtain negative values for the bound $B(\pi_{\exp(-\lambda_1 r)}, \gamma, \beta)$, and more generally $B(\rho, \gamma, \beta)$, leads the way to introducing the new concept of the *effective temperature* of an estimator.

Definition 2.1.1 For any posterior distribution $\rho : \Omega \rightarrow \mathcal{M}_+^1(\Theta)$ we define the *effective temperature* $T(\rho) \in \mathbb{R} \cup \{-\infty, +\infty\}$ of ρ by the equation

$$\rho(R) = \pi_{\exp(-\frac{R}{T(\rho)})}(R).$$

Note that $\beta \mapsto \pi_{\exp(-\beta R)}(R) : \mathbb{R} \cup \{-\infty, +\infty\} \rightarrow (0, 1)$ is continuous and strictly decreasing from $\operatorname{ess\,sup}_\pi R$ to $\operatorname{ess\,inf}_\pi R$ (as soon as these two bounds do not coincide). This shows that the effective temperature $T(\rho)$ is a well-defined random variable.

Theorem 2.1.3 provides a bound for $T(\rho)$, indeed:

Proposition 2.1.5. *Let*

$$\widehat{\beta}(\rho) = \sup\bigl\{\beta \in \mathbb{R}; \inf_{\gamma, N \tanh(\frac{\gamma}{N}) > \beta} B(\rho, \beta, \gamma) \leq 0\bigr\},$$

where $B(\rho, \beta, \gamma)$ is as in Theorem 2.1.3 (page 54). Then with $\mathbb{P}$ probability at least $1 - \epsilon$, for any posterior distribution $\rho : \Omega \rightarrow \mathcal{M}_+^1(\Theta)$, $T(\rho) \leq \widehat{\beta}(\rho)^{-1}$, or equivalently $\rho(R) \leq \pi_{\exp[-\widehat{\beta}(\rho) R]}(R)$.

This notion of *effective temperature* of a (randomized) estimator ρ is interesting for two reasons:

- the difference $\rho(R) - \pi_{\exp(-\beta R)}(R)$ can be estimated with better accuracy than $\rho(R)$ itself, due to the use of relative deviation inequalities, leading to convergence rates up to $1/N$ in favourable situations, even when $\inf_\Theta R$ is not close to zero;
- and of course $\pi_{\exp(-\beta R)}(R)$ is a decreasing function of β, thus being able to estimate $\rho(R) - \pi_{\exp(-\beta R)}(R)$ with some given accuracy, means being able to discriminate between values of $\rho(R)$ with the same accuracy, although doing so through the parametrization $\beta \mapsto \pi_{\exp(-\beta R)}(R)$, which can neither be observed nor estimated with the same precision!

2.1.3. Analysis of an empirical bound for the effective temperature. We are now going to launch into a mathematically rigorous analysis of the bound $B(\pi_{\exp(-\lambda_1 r), \beta, \gamma})$ provided by Theorem 2.1.3 (page 54), to show that $\inf_{\rho \in \mathcal{M}_+^1(\Theta)} \pi_{\exp[-\widehat{\beta}(\rho) R]}(R)$ converges indeed to $\inf_\Theta R$ at some optimal rate in favourable situations.

It is more convenient for this purpose to use deviation inequalities involving M' rather than m'. It is straightforward to extend Theorem 1.4.2 (page 35) to

Theorem 2.1.6. *For any real constants β and γ, for any prior distributions $\pi, \mu \in \mathcal{M}_+^1(\Theta)$, with $\mathbb{P}$ probability at least $1 - \eta$, for any posterior distribution $\rho : \Omega \rightarrow \mathcal{M}_+^1(\Theta)$,*

$$\gamma \rho \otimes \pi_{\exp(-\beta R)}\bigl[\Psi_{\frac{\gamma}{N}}(R', M')\bigr] \leq \gamma \rho \otimes \pi_{\exp(-\beta R)}(r') + \mathcal{K}(\rho, \mu) - \log(\eta).$$

In order to transform the left-hand side into a linear expression and in the same time localize this theorem, let us choose μ defined by its density

$$\frac{d\mu}{d\pi}(\theta_1) = C^{-1} \exp\biggl[-\beta R(\theta_1) - \gamma \int_\Theta \Bigl\{ \Psi_{\frac{\gamma}{N}}\bigl[R'(\theta_1,\theta_2), M'(\theta_1,\theta_2)\bigr] - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N}) R'(\theta_1,\theta_2)\Bigr\} \pi_{\exp(-\beta R)}(d\theta_2)\biggr],$$

where C is such that $\mu(\Theta) = 1$. We get

$$\begin{aligned}
\mathcal{K}(\rho,\mu) &= \beta\rho(R) + \gamma \rho \otimes \pi_{\exp(-\beta R)}\bigl[\Psi_{\frac{\gamma}{N}}(R',M') - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})R'\bigr] + \mathcal{K}(\rho,\pi) \\
&\quad + \log\biggl\{\int_\Theta \exp\biggl[-\beta R(\theta_1) - \gamma \int_\Theta \Bigl\{ \Psi_{\frac{\gamma}{N}}\bigl[R'(\theta_1,\theta_2), M'(\theta_1,\theta_2)\bigr] \\
&\qquad - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N}) R'(\theta_1,\theta_2)\Bigr\} \pi_{\exp(-\beta R)}(d\theta_2)\biggr] \pi(d\theta_1)\biggr\} \\
&= \beta\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] \\
&\quad + \gamma \rho \otimes \pi_{\exp(-\beta R)}\bigl[\Psi_{\frac{\gamma}{N}}(R',M') - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})R'\bigr] \\
&\quad + \mathcal{K}(\rho,\pi) - \mathcal{K}(\pi_{\exp(-\beta R)},\pi) \\
&\quad + \log\biggl\{\int_\Theta \exp\biggl[-\gamma \int_\Theta \Bigl\{ \Psi_{\frac{\gamma}{N}}\bigl[R'(\theta_1,\theta_2), M'(\theta_1,\theta_2)\bigr] \\
&\qquad - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N}) R'(\theta_1,\theta_2)\Bigr\} \pi_{\exp(-\beta R)}(d\theta_2)\biggr] \pi_{\exp(-\beta R)}(d\theta_1)\biggr\}.
\end{aligned}$$

Thus with $\mathbb{P}$ probability at least $1-\eta$,

$$\begin{aligned}
&\bigl[N\sinh(\tfrac{\gamma}{N}) - \beta\bigr]\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] \\
&\quad \leq \gamma\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] + \mathcal{K}(\rho,\pi) - \mathcal{K}(\pi_{\exp(-\beta R)},\pi) - \log(\eta) + C(\beta,\gamma) \\
&\quad \text{where } C(\beta,\gamma) = \log\biggl\{\int_\Theta \exp\biggl[-\gamma \int_\Theta \Bigl\{ \Psi_{\frac{\gamma}{N}}\bigl[R'(\theta_1,\theta_2), M'(\theta_1,\theta_2)\bigr] \\
&\qquad - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N}) R'(\theta_1,\theta_2)\Bigr\} \pi_{\exp(-\beta R)}(d\theta_2)\biggr] \pi_{\exp(-\beta R)}(d\theta_1)\biggr\}.
\end{aligned} \tag{2.6}$$

Remarking that

$$\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] = \beta\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] + \mathcal{K}(\rho,\pi) - \mathcal{K}(\pi_{\exp(-\beta R)},\pi),$$

we deduce from the previous inequality

THEOREM 2.1.7. *For any real constants β and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,*

$$N\sinh(\tfrac{\gamma}{N})\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] \leq \gamma\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] + \mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] - \log(\eta) + C(\beta,\gamma).$$

We can also go into a slightly different direction, starting back again from equation (2.6, page 57) and remarking that for any real constant λ,

$$\begin{aligned}\lambda\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] + \mathcal{K}(\rho, \pi) - \mathcal{K}(\pi_{\exp(-\beta R)}, \pi) \\ \leq \lambda\rho(r) + \mathcal{K}(\rho, \pi) + \log\bigl\{\pi\bigl[\exp(-\lambda r)\bigr]\bigr\} = \mathcal{K}\bigl[\rho, \pi_{\exp(-\lambda r)}\bigr].\end{aligned}$$

This leads to

THEOREM 2.1.8. *For any real constants β and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any real constant λ,*

$$\begin{aligned}\bigl[N\sinh(\tfrac{\gamma}{N}) - \beta\bigr]\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] \\ \leq (\gamma - \lambda)\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] + \mathcal{K}\bigl[\rho, \pi_{\exp(-\lambda r)}\bigr] - \log(\eta) + C(\beta, \gamma),\end{aligned}$$

where the definition of $C(\beta,\gamma)$ is given by equation (2.6, page 57).

We can now use this inequality in the case when $\rho = \pi_{\exp(-\lambda r)}$ and combine it with Inequality (2.5, page 53) to obtain

THEOREM 2.1.9 *For any real constants β and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any real constant λ,*

$$\bigl[\tfrac{N\lambda}{\beta}\sinh(\tfrac{\gamma}{N}) - \gamma\bigr]\bigl[\pi_{\exp(-\lambda r)}(r) - \pi_{\exp(-\beta R)}(r)\bigr] \leq C(\beta,\gamma) - \log(\eta).$$

We deduce from this theorem

PROPOSITION 2.1.10 *For any real positive constants β_1, β_2 and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any real constants λ_1 and λ_2, such that $\lambda_2 < \beta_2\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$ and $\lambda_1 > \beta_1\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$,*

$$\begin{aligned}\pi_{\exp(-\lambda_1 r)}(r) - \pi_{\exp(-\lambda_2 r)}(r) \leq \pi_{\exp(-\beta_1 R)}(r) - \pi_{\exp(-\beta_2 R)}(r) \\ + \frac{C(\beta_1,\gamma) + \log(2/\eta)}{\frac{N\lambda_1}{\beta_1}\sinh(\frac{\gamma}{N}) - \gamma} + \frac{C(\beta_2,\gamma) + \log(2/\eta)}{\gamma - \frac{N\lambda_2}{\beta_2}\sinh(\frac{\gamma}{N})}.\end{aligned}$$

Moreover, $\pi_{\exp(-\beta_1 R)}$ and $\pi_{\exp(-\beta_2 R)}$ being prior distributions, with $\mathbb{P}$ probability at least $1-\eta$,

$$\begin{aligned}\gamma\bigl[\pi_{\exp(-\beta_1 R)}(r) - \pi_{\exp(-\beta_2 R)}(r)\bigr] \\ \leq \gamma\pi_{\exp(-\beta_1 R)} \otimes \pi_{\exp(-\beta_2 R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R', M')\bigr] - \log(\eta).\end{aligned}$$

Hence

PROPOSITION 2.1.11 *For any positive real constants β_1, β_2 and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any positive real constants λ_1 and λ_2 such that $\lambda_2 < \beta_2\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$ and $\lambda_1 > \beta_1\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$,*

$$\begin{aligned}\pi_{\exp(-\lambda_1 r)}(r) - \pi_{\exp(-\lambda_2 r)}(r) \\ \leq \pi_{\exp(-\beta_1 R)} \otimes \pi_{\exp(-\beta_2 R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R', M')\bigr] \\ + \frac{\log(\frac{3}{\eta})}{\gamma} + \frac{C(\beta_1,\gamma) + \log(\frac{3}{\eta})}{\frac{N\lambda_1}{\beta_1}\sinh(\frac{\gamma}{N}) - \gamma} + \frac{C(\beta_2,\gamma) + \log(\frac{3}{\eta})}{\gamma - \frac{N\lambda_2}{\beta_2}\sinh(\frac{\gamma}{N})}.\end{aligned}$$

In order to achieve the analysis of the bound $B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma)$ given by Theorem 2.1.3 (page 54), it now remains to bound quantities of the general form

$$\log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda r)}(m')\bigr\}\Bigr]\Bigr\} \\ = \sup_{\rho\in\mathcal{M}_+^1(\Theta)} N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho\otimes\pi_{\exp(-\lambda)}(m') - \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda r)}\bigr].$$

Let us consider the prior distribution $\mu\in\mathcal{M}_+^1(\Theta\times\Theta)$ on couples of parameters defined by the density

$$\frac{d\mu}{d(\pi\otimes\pi)}(\theta_1,\theta_2) = C^{-1}\exp\Bigl\{-\beta R(\theta_1) - \beta R(\theta_2) + \alpha\Phi_{-\frac{\alpha}{N}}\bigl[M'(\theta_1,\theta_2)\bigr]\Bigr\},$$

where the normalizing constant C is such that $\mu(\Theta\times\Theta)=1$. Since for fixed values of the parameters θ and $\theta'\in\Theta$, $m'(\theta,\theta')$, like $r(\theta)$, is a sum of independent Bernoulli random variables, we can easily adapt the proof of Theorem 1.1.4 on page 4, to establish that with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distribution ρ and any real constant λ,

$$\alpha\rho\otimes\pi_{\exp(-\lambda r)}(m') \leq \alpha\rho\otimes\pi_{\exp(-\lambda r)}\bigl[\Phi_{-\frac{\alpha}{N}}(M')\bigr] \\ + \mathcal{K}(\rho\otimes\pi_{\exp(-\lambda r)},\mu) - \log(\eta) \\ = \mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr] + \mathcal{K}\bigl[\pi_{\exp(-\lambda r)},\pi_{\exp(-\beta R)}\bigr] \\ + \log\Bigl\{\pi_{\exp(-\beta R)}\otimes\pi_{\exp(-\beta R)}\Bigl[\exp\bigl(\alpha\Phi_{-\frac{\alpha}{N}}\circ M'\bigr)\Bigr]\Bigr\} - \log(\eta).$$

Thus for any real constant β and any positive real constants α and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any real constant λ,

$$(2.7)\quad \log\Bigl\{\pi_{\exp(-\lambda r)}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi_{\exp(-\lambda r)}(m')\bigr\}\Bigr]\Bigr\} \\ \leq \sup_{\rho\in\mathcal{M}_+^1(\Theta)}\biggl(\tfrac{N}{\alpha}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\Bigl\{\mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr] + \mathcal{K}\bigl[\pi_{\exp(-\lambda r)},\pi_{\exp(-\beta R)}\bigr] \\ + \log\bigl\{\pi_{\exp(-\beta R)}\otimes\pi_{\exp(-\beta R)}\bigl[\exp(\alpha\Phi_{-\frac{\alpha}{N}}\circ M')\bigr]\bigr\} \\ - \log(\eta)\Bigr\} - \mathcal{K}\bigl[\rho,\pi_{\exp(-\lambda r)}\bigr]\biggr).$$

To finish, we need some appropriate upper bound for the entropy $\mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr]$. This question can be handled in the following way: using Theorem 2.1.7 (page 57), we see that for any positive real constants γ and β, with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distribution ρ,

$$\mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr] = \beta\bigl[\rho(R) - \pi_{\exp(-\beta R)}(R)\bigr] + \mathcal{K}(\rho,\pi) - \mathcal{K}(\pi_{\exp(-\beta R)},\pi) \\ \leq \frac{\beta}{N\sinh(\frac{\gamma}{N})}\biggl[\gamma\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] \\ + \mathcal{K}\bigl[\rho,\pi_{\exp(-\beta R)}\bigr] - \log(\eta) + C(\beta,\gamma)\biggr] \\ + \mathcal{K}(\rho,\pi) - \mathcal{K}(\pi_{\exp(-\beta R)},\pi) \\ \leq \mathcal{K}\bigl[\rho,\pi_{\exp(-\frac{\beta\gamma}{N\sinh(\frac{\gamma}{N})}r)}\bigr]$$

$$+ \frac{\beta}{N \sinh(\frac{\gamma}{N})} \Bigl\{ \mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] + C(\beta, \gamma) - \log(\eta) \Bigr\}.$$

In other words,

THEOREM 2.1.12. *For any positive real constants β and γ such that $\beta < N \times \sinh(\frac{\gamma}{N})$, with $\mathbb{P}$ probability at least $1 - \eta$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,*

$$\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] \leq \frac{\mathcal{K}\bigl[\rho, \pi_{\exp[-\beta \frac{\gamma}{N} \sinh(\frac{\gamma}{N})^{-1} r]}\bigr]}{1 - \frac{\beta}{N \sinh(\frac{\gamma}{N})}} + \frac{C(\beta, \gamma) - \log(\eta)}{\frac{N \sinh(\frac{\gamma}{N})}{\beta} - 1},$$

where the quantity $C(\beta, \gamma)$ is defined by equation (2.6, page 57). Equivalently, it will be in some cases more convenient to use this result in the form: for any positive real constants λ and γ, with $\mathbb{P}$ probability at least $1 - \eta$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,

$$\mathcal{K}\bigl[\rho, \pi_{\exp[-\lambda \frac{N}{\gamma} \sinh(\frac{\gamma}{N}) R]}\bigr] \leq \frac{\mathcal{K}\bigl[\rho, \pi_{\exp(-\lambda r)}\bigr]}{1 - \frac{\lambda}{\gamma}} + \frac{C(\lambda \frac{N}{\gamma} \sinh(\frac{\gamma}{N}), \gamma) - \log(\eta)}{\frac{\lambda}{\beta} - 1}.$$

Choosing in equation (2.7, page 59) $\alpha = \dfrac{N \log\bigl[\cosh(\frac{\gamma}{N})\bigr]}{1 - \frac{\beta}{N \sinh(\frac{\gamma}{N})}}$ and $\beta = \lambda \frac{N}{\gamma} \sinh(\frac{\gamma}{N})$, so that $\alpha = \dfrac{N \log\bigl[\cosh(\frac{\gamma}{N})\bigr]}{1 - \frac{\lambda}{\gamma}}$, we obtain with $\mathbb{P}$ probability at least $1 - \eta$,

$$\begin{aligned}
&\log\Bigl\{ \pi_{\exp(-\lambda r)} \Bigl[\exp\bigl\{ N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] \pi_{\exp(-\lambda r)}(m') \bigr\} \Bigr] \Bigr\} \\
&\quad \leq \tfrac{2\lambda}{\gamma} \bigl[C(\beta, \gamma) + \log(\tfrac{2}{\eta}) \bigr] \\
&\qquad + \Bigl(1 - \tfrac{\lambda}{\gamma}\Bigr) \Bigl[\log\Bigl\{ \pi_{\exp(-\beta R)} \otimes \pi_{\exp(-\beta R)} \bigl[\exp(\alpha \Phi_{-\frac{\alpha}{N}} \circ M') \bigr] \Bigr\} \\
&\qquad\qquad + \log(\tfrac{2}{\eta}) \Bigr].
\end{aligned}$$

This proves

PROPOSITION 2.1.13. *For any positive real constants $\lambda < \gamma$, with $\mathbb{P}$ probability at least $1 - \eta$,*

$$\begin{aligned}
&\log\Bigl\{ \pi_{\exp(-\lambda r)} \Bigl[\exp\bigl\{ N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] \pi_{\exp(-\lambda r)}(m') \bigr\} \Bigr] \Bigr\} \\
&\quad \leq \frac{2\lambda}{\gamma} \bigl[C(\tfrac{N\lambda}{\gamma} \sinh(\tfrac{\gamma}{N}), \gamma) + \log(\tfrac{2}{\eta}) \bigr] \\
&\qquad + \Bigl(1 - \tfrac{\lambda}{\gamma}\Bigr) \log\Bigl\{ \pi^{\otimes 2}_{\exp[-\frac{N\lambda}{\gamma} \sinh(\frac{\gamma}{N}) R]} \Bigl[\\
&\qquad\qquad \exp\Bigl(\frac{N \log[\cosh(\frac{\gamma}{N})]}{1 - \frac{\lambda}{\gamma}} \Phi_{-\frac{\log[\cosh(\frac{\gamma}{N})]}{1 - \frac{\lambda}{\gamma}}} \circ M' \Bigr) \Bigr] \Bigr\} \\
&\qquad\qquad + \Bigl(1 - \tfrac{\lambda}{\gamma}\Bigr) \log(\tfrac{2}{\eta}).
\end{aligned}$$

We are now ready to analyse the bound $B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma)$ of Theorem 2.1.3 (page 54).

THEOREM 2.1.14. *For any positive real constants λ_1, λ_2, β_1, β_2, β and γ, such that*

$$\begin{aligned} &\lambda_1 < \gamma, & &\beta_1 < \tfrac{N\lambda_1}{\gamma}\sinh(\tfrac{\gamma}{N}),\\ &\lambda_2 < \gamma, & &\beta_2 > \tfrac{N\lambda_2}{\gamma}\sinh(\tfrac{\gamma}{N}),\\ & & &\beta < \tfrac{N\lambda_2}{\gamma}\tanh(\tfrac{\gamma}{N}), \end{aligned}$$

with $\mathbb{P}$ probability $1-\eta$, the bound $B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma)$ of Theorem 2.1.3 (page 54) satisfies

$$\begin{aligned} &B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma)\\ &\quad\le (\gamma-\lambda_1)\biggl\{\pi_{\exp(-\beta_1 R)}\otimes\pi_{\exp(-\beta_2 R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R', M')\bigr] + \frac{\log(\frac{7}{\eta})}{\gamma}\\ &\qquad + \frac{C(\beta_1,\gamma)+\log(\frac{7}{\eta})}{\frac{N\lambda_1}{\beta_1}\sinh(\frac{\gamma}{N})-\gamma} + \frac{C(\beta_2,\gamma)+\log(\frac{7}{\eta})}{\gamma-\frac{N\lambda_2}{\beta_2}\sinh(\frac{\gamma}{N})}\biggr\}\\ &\qquad + \frac{2\lambda_1}{\gamma}\Bigl[C\bigl(\tfrac{N\lambda_1}{\gamma}\sinh(\tfrac{\gamma}{N}),\gamma\bigr)+\log(\tfrac{7}{\eta})\Bigr]\\ &\quad + \Bigl(1-\tfrac{\lambda_1}{\gamma}\Bigr)\log\biggl\{\pi^{\otimes 2}_{\exp[-\frac{N\lambda_1}{\gamma}\sinh(\frac{\gamma}{N})R]}\biggl[\\ &\qquad \exp\biggl(\frac{N\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_1}{\gamma}}\Phi_{-\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_1}{\gamma}}}\circ M'\biggr)\biggr]\biggr\}\\ &\qquad + \Bigl(1-\tfrac{\lambda_1}{\gamma}\Bigr)\log(\tfrac{7}{\eta}) - \log\bigl[\nu(\{\beta\})\nu(\{\gamma\})\epsilon\bigr]\\ &\quad + (\gamma-\lambda_1)\tfrac{\beta}{\lambda_2}F^{-1}_{\gamma,\frac{\beta\gamma}{\lambda_2}}\biggl\{\frac{2\lambda_2}{\gamma}\Bigl[C\bigl(\tfrac{N\lambda_2}{\gamma}\sinh(\tfrac{\gamma}{N}),\gamma\bigr)+\log\bigl(\tfrac{7}{\eta}\bigr)\Bigr]\\ &\qquad + \Bigl(1-\tfrac{\lambda_2}{\gamma}\Bigr)\log\biggl\{\pi^{\otimes 2}_{\exp[-\frac{N\lambda_2}{\gamma}\sinh(\frac{\gamma}{N})R]}\biggl[\\ &\qquad \exp\biggl(\frac{N\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_2}{\gamma}}\Phi_{-\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_2}{\gamma}}}\circ M'\biggr)\biggr]\biggr\}\\ &\qquad + \Bigl(1-\tfrac{\lambda_2}{\gamma}\Bigr)\log\bigl(\tfrac{7}{\eta}\bigr) - \log\bigl[\nu(\{\beta\})\nu(\{\gamma\})\epsilon\bigr]\biggr\}, \end{aligned}$$

where the function $C(\beta,\gamma)$ is defined by equation (2.6, page 57).

2.1.4. ADAPTATION TO PARAMETRIC AND MARGIN ASSUMPTIONS. To help understand the previous theorem, it may be useful to give linear upper-bounds to the factors appearing in the right-hand side of the previous inequality. Introducing $\widetilde{\theta}$ such that $R(\widetilde{\theta}) = \inf_\Theta R$ (assuming that such a parameter exists) and remembering that

$$\begin{aligned} \Psi_{-a}(p,m) &\le a^{-1}\sinh(a)p + 2a^{-1}\sinh(\tfrac{a}{2})^2 m, & a\in\mathbb{R}_+,\\ \Phi_{-a}(p) &\le a^{-1}\bigl[\exp(a)-1\bigr]p, & a\in\mathbb{R}_+, \end{aligned}$$

$$\begin{aligned}
\Psi_a(p,m) &\geq a^{-1}\sinh(a)p - 2a^{-1}\sinh(\tfrac{a}{2})^2 m, && a\in\mathbb{R}_+,\\
M'(\theta_1,\theta_2) &\leq M'(\theta_1,\widetilde{\theta}) + M'(\theta_2,\widetilde{\theta}), && \theta_1,\theta_2\in\Theta,\\
M'(\theta_1,\widetilde{\theta}) &\leq xR'(\theta_1,\widetilde{\theta}) + \varphi(x), && x\in\mathbb{R}_+,\theta_1\in\Theta,
\end{aligned}$$

the last inequality being rather a consequence of the definition of φ than a property of M', we easily see that

$$\begin{aligned}
&\pi_{\exp(-\beta_1 R)}\otimes\pi_{\exp(-\beta_2 R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R',M')\bigr]\\
&\quad\leq \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})\bigl[\pi_{\exp(-\beta_1 R)}(R) - \pi_{\exp(-\beta_2 R)}(R)\bigr]\\
&\qquad\qquad + \tfrac{2N}{\gamma}\sinh(\tfrac{\gamma}{2N})^2\pi_{\exp(-\beta_1 R)}\otimes\pi_{\exp(-\beta_2 R)}(M')\\
&\quad\leq \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})\bigl[\pi_{\exp(-\beta_1 R)}(R) - \pi_{\exp(-\beta_2 R)}(R)\bigr]\\
&\qquad + \frac{2xN}{\gamma}\sinh(\tfrac{\gamma}{2N})^2\Bigl\{\pi_{\exp(-\beta_1 R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr] + \pi_{\exp(-\beta_2 R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr]\Bigr\}\\
&\qquad\qquad + \frac{4N}{\gamma}\sinh(\tfrac{\gamma}{2N})^2\varphi(x),
\end{aligned}$$

that

$$\begin{aligned}
C(\beta,\gamma) &\leq \log\Bigl\{\pi_{\exp(-\beta R)}\Bigl\{\exp\Bigl[2N\sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2\pi_{\exp(-\beta R)}(M')\Bigr]\Bigr\}\Bigr\}\\
&\leq \log\Bigl\{\pi_{\exp(-\beta R)}\Bigl\{\exp\Bigl[2N\sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 M'(\cdot,\widetilde{\theta})\Bigr]\Bigr\}\Bigr\}\\
&\qquad + 2N\sinh(\tfrac{\gamma}{2N})^2\pi_{\exp(-\beta R)}\bigl[M'(\cdot,\widetilde{\theta})\bigr]\\
&\leq \log\Bigl\{\pi_{\exp(-\beta R)}\Bigl\{\exp\Bigl[2xN\sinh(\tfrac{\gamma}{2N})^2 R'(\cdot,\widetilde{\theta})\Bigr]\Bigr\}\Bigr\}\\
&\qquad + 2xN\sinh(\tfrac{\gamma}{2N})^2\pi_{\exp(-\beta R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr] + 4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x)\\
&= \int_{\beta-2xN\sinh(\frac{\gamma}{2N})^2}^{\beta}\pi_{\exp(-\alpha R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr]d\alpha\\
&\qquad + 2xN\sinh(\tfrac{\gamma}{2N})^2\pi_{\exp(-\beta R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr] + 4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x)\\
&\leq 4xN\sinh(\tfrac{\gamma}{2N})^2\pi_{\exp[-(\beta-2xN\sinh(\frac{\gamma}{2N})^2)R]}\bigl[R'(\cdot,\widetilde{\theta})\bigr]\\
&\qquad + 4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x),
\end{aligned}$$

and that

$$\begin{aligned}
&\log\Bigl\{\pi_{\exp(-\beta R)}^{\otimes 2}\Bigl[\exp\Bigl(N\alpha\Phi_{-\alpha}\circ M'\Bigr)\Bigr]\Bigr\}\\
&\qquad\leq 2\log\Bigl\{\pi_{\exp(-\beta R)}\Bigl[\exp\Bigl(N\bigl[\exp(\alpha)-1\bigr]M'(\cdot,\widetilde{\theta})\Bigr)\Bigr]\Bigr\}\\
&\qquad\leq 2xN\bigl[\exp(\alpha)-1\bigr]\pi_{\exp[-(\beta-xN[\exp(\alpha)-1])R]}\bigl[R'(\cdot,\widetilde{\theta})\bigr]\\
&\qquad\qquad + 2xN\bigl[\exp(\alpha)-1\bigr]\varphi(x).
\end{aligned}$$

Let us push further the investigation under the parametric assumption that for some positive real constant d

$$\lim_{\beta\to+\infty}\beta\pi_{\exp(-\beta R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr] = d, \tag{2.8}$$

This assumption will for instance hold true with $d = \frac{n}{2}$ when $R : \Theta \to (0,1)$ is a smooth function defined on a compact subset Θ of $\mathbb{R}^n$ that reaches its minimum value on a finite number of non-degenerate (i.e. with a positive definite Hessian) interior points of Θ, and π is absolutely continuous with respect to the Lebesgue measure on Θ and has a smooth density.

In case of assumption (2.8), if we restrict ourselves to sufficiently large values of the constants β, β_1, β_2, λ_1, λ_2 and γ (the smaller of which is as a rule β, as we will see), we can use the fact that for some (small) positive constant δ, and some (large) positive constant A,

$$\frac{d}{\alpha}(1-\delta) \leq \pi_{\exp(-\alpha R)}\bigl[R'(\cdot,\widetilde{\theta})\bigr] \leq \frac{d}{\alpha}(1+\delta), \qquad \alpha \geq A. \tag{2.9}$$

Under this assumption,

$$\begin{aligned}
&\pi_{\exp(-\beta_1 R)} \otimes \pi_{\exp(-\beta_2 R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R',M')\bigr] \\
&\quad \leq \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})\bigl[\tfrac{d}{\beta_1}(1+\delta) - \tfrac{d}{\beta_2}(1-\delta)\bigr] \\
&\qquad + \tfrac{2xN}{\gamma}\sinh(\tfrac{\gamma}{2N})^2(1+\delta)\bigl[\tfrac{d}{\beta_1}+\tfrac{d}{\beta_2}\bigr] + \tfrac{4N}{\gamma}\sinh(\tfrac{\gamma}{2N})^2\varphi(x).
\end{aligned}$$

$$\begin{aligned}
C(\beta,\gamma) &\leq d(1+\delta)\log\Bigl(\tfrac{\beta}{\beta - 2xN\sinh(\frac{\gamma}{2N})^2}\Bigr) \\
&\qquad + 2xN\sinh(\tfrac{\gamma}{2N})^2\tfrac{(1+\delta)d}{\beta} + 4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x).
\end{aligned}$$

$$\begin{aligned}
&\log\Bigl\{\pi_{\exp(-\beta R)}^{\otimes 2}\Bigl[\exp\Bigl(N\alpha\Phi_{-\alpha}\circ M'\Bigr)\Bigr]\Bigr\} \\
&\quad \leq 2xN\bigl[\exp(\alpha)-1\bigr]\frac{d(1+\delta)}{\beta - xN[\exp(\alpha)-1]} + 2N\bigl[\exp(\alpha)-1\bigr]\varphi(x).
\end{aligned}$$

Thus with $\mathbb{P}$ probability at least $1-\eta$,

$$\begin{aligned}
&B(\pi_{\exp(-\lambda_1 r)},\beta,\gamma) \leq -(\gamma-\lambda_1)\tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})\tfrac{d}{\beta_2}(1-\delta) \\
&+ (\gamma-\lambda_1)\Biggl\{\tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N})\tfrac{(1+\delta)d}{\beta_1} \\
&\qquad + \tfrac{2xN}{\gamma}\sinh(\tfrac{\gamma}{2N})^2(1+\delta)\bigl[\tfrac{d}{\beta_1}+\tfrac{d}{\beta_2}\bigr] + \tfrac{4N}{\gamma}\sinh(\tfrac{\gamma}{2N})^2\varphi(x) + \frac{\log(\frac{7}{\eta})}{\gamma} \\
&\quad + \frac{4xN\sinh(\frac{\gamma}{2N})^2\frac{(1+\delta)d}{\beta_1 - 2xN\sinh(\frac{\gamma}{2N})^2} + 4N\sinh(\frac{\gamma}{2N})^2\varphi(x) + \log(\frac{7}{\eta})}{\frac{N\lambda_1}{\beta_1}\sinh(\frac{\gamma}{N}) - \gamma} \\
&\qquad + \frac{4xN\sinh(\frac{\gamma}{2N})^2\frac{(1+\delta)d}{\beta_2 - 2xN\sinh(\frac{\gamma}{2N})^2} + 4N\sinh(\frac{\gamma}{2N})^2\varphi(x) + \log(\frac{7}{\eta})}{\gamma - \frac{N\lambda_2}{\beta_2}\sinh(\frac{\gamma}{N})}\Biggr\} \\
&+ \frac{2\lambda_1}{\gamma}\Biggl\{4xN\sinh(\tfrac{\gamma}{2N})^2\frac{(1+\delta)d}{\frac{N\lambda_1}{\gamma}\sinh(\frac{\gamma}{N}) - 2xN\sinh(\frac{\gamma}{2N})^2} \\
&\qquad\qquad + 4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x) + \log(\tfrac{7}{\eta})\Biggr\} \\
&+ \Bigl(1-\frac{\lambda_1}{\gamma}\Bigr)\Biggl\{2d(1+\delta)\Biggl(\frac{\lambda_1\sinh\bigl(\frac{\gamma}{N}\bigr)}{x\gamma\Bigl[\exp\Bigl(\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_1}{\gamma}}\Bigr)-1\Bigr]} - 1\Biggr)^{-1} \\
&\qquad\qquad + 2N\Bigl[\exp\Bigl(\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_1}{\gamma}}\Bigr) - 1\Bigr]\varphi(x)\Biggr\}
\end{aligned}$$

$$
\begin{aligned}
&+\Big(1-\tfrac{\lambda_1}{\gamma}\Big)\log(\tfrac{7}{\eta})-\log\big[\nu(\{\beta\})\nu(\{\gamma\})\epsilon\big]\\
&+\frac{1-\frac{\lambda_1}{\gamma}}{\frac{N\lambda_2}{\beta\gamma}\tanh(\frac{\gamma}{N})-1}\Bigg\{\frac{2\lambda_2}{\gamma}\bigg\{4xN\sinh(\tfrac{\gamma}{2N})^2\frac{(1+\delta)d}{\frac{N\lambda_2}{\gamma}\sinh(\frac{\gamma}{N})-2xN\sinh(\frac{\gamma}{2N})^2}\\
&\qquad+4N\sinh(\tfrac{\gamma}{2N})^2\varphi(x)+\log(\tfrac{7}{\eta})\bigg\}\\
&+\Big(1-\frac{\lambda_2}{\gamma}\Big)\Bigg[2d(1+\delta)\Bigg(\frac{\lambda_2\sinh\big(\frac{\gamma}{N}\big)}{x\gamma\Big[\exp\Big(\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_2}{\gamma}}\Big)-1\Big]}-1\Bigg)^{-1}\\
&\qquad+2N\Big[\exp\Big(\frac{\log[\cosh(\frac{\gamma}{N})]}{1-\frac{\lambda_2}{\gamma}}\Big)-1\Big]\varphi(x)\Bigg]\\
&\qquad+\Big(1-\tfrac{\lambda_2}{\gamma}\Big)\log(\tfrac{7}{\eta})-\log\big[\nu(\beta)\nu(\gamma)\epsilon\big]\Bigg\}.
\end{aligned}
$$

Now let us choose for simplicity $\beta_2 = 2\lambda_2 = 4\beta$, $\beta_1 = \lambda_1/2 = \gamma/4$, and let us introduce the notation

$$
\begin{aligned}
C_1 &= \frac{N}{\gamma}\sinh(\frac{\gamma}{N}),\\
C_2 &= \frac{N}{\gamma}\tanh(\frac{\gamma}{N}),\\
C_3 &= \frac{N^2}{\gamma^2}\big[\exp(\frac{\gamma^2}{N^2})-1\big]\\
\text{and}\quad C_4 &= \frac{2N^2(1-\frac{2\beta}{\gamma})}{\gamma^2}\Big[\exp\Big(\frac{\gamma^2}{2N^2(1-\frac{2\beta}{\gamma})}\Big)-1\Big],
\end{aligned}
$$

to obtain

$$
\begin{aligned}
B(\pi_{\exp(-\lambda_1 r)},\beta,\gamma) &\le -\frac{C_1\gamma}{8\beta}(1-\delta)d\\
&+\frac{C_1\gamma}{2}\Big\{\tfrac{4(1+\delta)d}{\gamma}+x\tfrac{\gamma}{2N}(1+\delta)\big[\tfrac{4d}{\gamma}+\tfrac{d}{4\beta}\big]+\tfrac{\gamma}{N}\varphi(x)\Big\}+\tfrac12\log(\tfrac{7}{\eta})\\
&\quad+\frac{1}{2C_1-1}\Big[(1+\delta)d\Big(\tfrac{N}{2xC_1\gamma}-1\Big)^{-1}+C_1\frac{\gamma^2}{2N}\varphi(x)+\tfrac12\log(\tfrac{7}{\eta})\Big]\\
&\qquad+\frac{1}{2-C_1}\Big[2(1+\delta)d\Big(\tfrac{8N\beta}{xC_1\gamma^2}-1\Big)^{-1}+C_1\frac{\gamma^2}{N}\varphi(x)+\log(\tfrac{7}{\eta})\Big]\\
&\qquad\qquad+\frac{2x\gamma(1+\delta)d}{N-x\gamma}+C_1\tfrac{\gamma^2}{N}\varphi(x)+\log(\tfrac{7}{\eta})\\
&\quad+d(1+\delta)\frac{x\gamma}{N}\Big(\frac{C_1}{2C_3}-\frac{x\gamma}{N}\Big)^{-1}+\frac{\gamma^2}{N}C_3\varphi(x)+\frac{\log(\frac{7}{\eta})}{2}-\log\big[\nu(\beta)\nu(\gamma)\epsilon\big]\\
&+\Big(4C_2-2\Big)^{-1}\Bigg\{\frac{4\beta}{\gamma}\bigg\{x\frac{\gamma^2}{N}C_1(1+\delta)d\Big(2\beta C_1-xC_1\frac{\gamma^2}{2N}\Big)^{-1}\\
&\qquad+\tfrac{\gamma^2}{N}\varphi(x)+\log(\tfrac{7}{\eta})\bigg\}\\
&+\Big(1-\frac{2\beta}{\gamma}\Big)\bigg\{2d(1+\delta)\frac{x\gamma}{N}\Big[\frac{4\beta C_1}{\gamma C_4}\Big(1-\frac{2\beta}{\gamma}\Big)-\frac{x\gamma}{N}\Big]^{-1}
\end{aligned}
$$

$$+ \frac{\gamma^2}{N(1-\frac{2\beta}{\gamma})} C_4 \varphi(x) \Big\}$$

$$+ \Big(1 - \tfrac{2\beta}{\gamma}\Big) \log(\tfrac{7}{\eta}) - \log\big[\nu(\beta)\nu(\gamma)\epsilon\big] \Big\}.$$

This simplifies to

$$\begin{aligned} B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma) \leq & -\frac{C_1}{8}(1-\delta)d\frac{\gamma}{\beta} \\ & + 2C_1(1+\delta)d + \log(\tfrac{7}{\eta})\Big[2 + \tfrac{3C_1}{(4C_1-2)(2-C_1)} + \frac{1+\frac{2\beta}{\gamma}}{4C_2-2}\Big] \\ & \qquad - \big(1 + \tfrac{1}{4C_2-2}\big)\log\big[\nu(\beta)\nu(\gamma)\epsilon\big] \\ & + \frac{(1+\delta)dx\gamma}{N}\Big\{ C_1 + \tfrac{1}{2C_1-1}\Big(\tfrac{1}{2C_1} - \tfrac{\gamma x}{N}\Big)^{-1} \\ & \qquad + 2\Big(1 - \tfrac{\gamma x}{N}\Big)^{-1} + \Big(\tfrac{C_1}{2C_3} - \tfrac{\gamma x}{N}\Big)^{-1} + \tfrac{4C_1\beta}{\gamma(4C_2-2)}\Big\} \\ & + \frac{(1+\delta)dx\gamma^2}{N\beta}\Big\{\tfrac{C_1}{16} + \tfrac{2}{2-C_1}\Big(\tfrac{8}{C_1} - \tfrac{x\gamma^2}{N\beta}\Big)^{-1} \\ & \qquad + \Big(1 - \tfrac{2\beta}{\gamma}\Big)\tfrac{1}{2C_2-1}\Big[\tfrac{4C_1}{C_4}\Big(1 - \tfrac{2\beta}{\gamma}\Big) - \tfrac{\gamma^2 x}{\beta N}\Big]^{-1}\Big\} \\ & + \frac{\gamma^2}{N}\varphi(x)\Big\{\tfrac{3C_1}{2} + \tfrac{C_1}{4C_1-2} + \tfrac{C_1}{2-C_1} + C_3 + \tfrac{4\beta}{\gamma(4C_2-2)} + \tfrac{C_4}{4C_2-2}\Big\}. \end{aligned}$$

This shows that there exist universal positive real constants A_1, A_2, B_1, B_2, B_3, and B_4 such that as soon as $\frac{\gamma \max\{x,1\}}{N} \leq A_1 \frac{\beta}{\gamma} \leq A_2$,

$$\begin{aligned} B(\pi_{\exp(-\lambda_1 r)}, \beta, \gamma) \leq -B_1(1-\delta)d\frac{\gamma}{\beta} + B_2(1+\delta)d \\ - B_3 \log\big[\nu(\beta)\nu(\gamma)\epsilon\,\eta\big] + B_4\frac{\gamma^2}{N}\varphi(x). \end{aligned}$$

Thus $\pi_{\exp(-\lambda_1 r)}(R) \leq \pi_{\exp(-\beta R)}(R) \leq \inf_\Theta R + \frac{(1+\delta)d}{\beta}$ as soon as

$$\frac{\beta}{\gamma} \leq \frac{B_1}{B_2\frac{(1+\delta)}{(1-\delta)} + \frac{B_4\frac{\gamma^2}{N}\varphi(x) - B_3\log[\nu(\beta)\nu(\gamma)\epsilon\eta]}{(1-\delta)d}}.$$

Choosing some real ratio $\alpha > 1$, we can now make the above result uniform for any

$$\beta, \gamma \in \Lambda_\alpha \overset{\text{def}}{=} \Big\{\alpha^k ; k \in \mathbb{N}, 0 \leq k < \tfrac{\log(N)}{\log(\alpha)}\Big\}, \tag{2.10}$$

by substituting $\nu(\beta)$ and $\nu(\gamma)$ with $\frac{\log(\alpha)}{\log(\alpha N)}$ and $-\log(\eta)$ with $-\log(\eta) + 2 \times \log\Big[\frac{\log(\alpha N)}{\log(\alpha)}\Big]$.

Taking $\eta = \epsilon$ for simplicity, we can summarize our result in

THEOREM 2.1.15. *There exist positive real universal constants A, B_1, B_2, B_3 and B_4 such that for any positive real constants $\alpha > 1$, d and δ, for any prior* *distribution* $\pi \in \mathcal{M}_+^1(\Theta)$, *with* $\mathbb{P}$ *probability at least* $1-\epsilon$, *for any* $\beta, \gamma \in \Lambda_\alpha$ *(where* Λ_α *is defined by equation* (2.10) *above) such that*

$$\sup_{\beta' \in \mathbb{R}, \beta' \geq \beta} \left| \frac{\beta'}{d} \bigl[\pi_{\exp(-\beta' R)}(R) - \inf_\Theta R \bigr] - 1 \right| \leq \delta$$

and such that also for some positive real parameter x

$$\frac{\gamma \max\{x, 1\}}{N} \leq \frac{A\beta}{\gamma} \text{ and } \frac{\beta}{\gamma} \leq \frac{B_1}{B_2 \frac{(1+\delta)}{(1-\delta)} + \frac{B_4 \frac{\gamma^2}{N} \varphi(x) - 2B_3 \log(\epsilon) + 4B_3 \log\left[\frac{\log(N)}{\log(\alpha)}\right]}{(1-\delta)d}},$$

the bound $B(\pi_{\exp(-\frac{\gamma}{2}r)}, \beta, \gamma)$ *given by Theorem 2.1.3 on page 54 in the case where we have chosen* ν *to be the uniform probability measure on* Λ_α, *satisfies* $B(\pi_{\exp(-\frac{\gamma}{2}r)}, \beta, \gamma) \leq 0$, *proving that* $\widehat{\beta}(\pi_{\exp(-\frac{\gamma}{2}r)}) \geq \beta$ *and therefore that*

$$\pi_{\exp(-\gamma \frac{r}{2})}(R) \leq \pi_{\exp(-\beta R)}(R) \leq \inf_\Theta R + \frac{(1+\delta)d}{\beta}.$$

What is important in this result is that we do not only bound $\pi_{\exp(-\frac{\gamma}{2}r)}(R)$, but also $B(\pi_{\exp(-\frac{\gamma}{2}r)}, \beta, \gamma)$, and that we do it uniformly on a grid of values of β and γ, showing that we can indeed set the constants β and γ adaptively using the empirical bound $B(\pi_{\exp(-\frac{\gamma}{2}r)}, \beta, \gamma)$.

Let us see what we get under the margin assumption (1.24, page 39). When $\kappa = 1$, we have $\varphi(c^{-1}) \leq 0$, leading to

Corollary 2.1.16. *Assuming that the margin assumption (1.24, page 39) is satisfied for* $\kappa = 1$, *that* $R : \Theta \to (0,1)$ *is independent of* N *(which is the case for instance when* $\mathbb{P} = P^{\otimes N}$*), and is such that*

$$\lim_{\beta' \to +\infty} \beta' \bigl[\pi_{\exp(-\beta' R)}(R) - \inf_\Theta R \bigr] = d,$$

there are universal positive real constants B_5 *and* B_6 *and* $N_1 \in \mathbb{N}$ *such that for any* $N \geq N_1$, *with* $\mathbb{P}$ *probability at least* $1-\epsilon$

$$\pi_{\exp(-\widehat{\gamma}\frac{r}{2})}(R) \leq \inf_\Theta R + \frac{B_5 d}{cN} \left[1 + \frac{B_6}{d} \log\left(\frac{\log(N)}{\epsilon} \right) \right]^2,$$

where $\widehat{\gamma} \in \arg\max_{\gamma \in \Lambda_2} \max\bigl\{ \beta \in \Lambda_2 ; B(\pi_{\exp(-\gamma \frac{r}{2})}, \beta, \gamma) \leq 0 \bigr\}$, *where* Λ_2 *is defined by equation (2.10, page 65), and* B *is the bound of Theorem 2.1.3 (page 54).*

When $\kappa > 1$, $\varphi(x) \leq (1-\kappa^{-1})\bigl(\kappa c x\bigr)^{-\frac{1}{\kappa-1}}$, and we can choose γ and x such that $\frac{\gamma^2}{N}\varphi(x) \simeq d$ to prove

Corollary 2.1.17. *Assuming that the margin assumption (1.24, page 39) is satisfied for some exponent* $\kappa > 1$, *that* $R : \Theta \to (0,1)$ *is independent of* N *(which is for instance the case when* $\mathbb{P} = P^{\otimes N}$*), and is such that*

$$\lim_{\beta' \to +\infty} \beta' \bigl[\pi_{\exp(-\beta' R)}(R) - \inf_\Theta R \bigr] = d,$$

there are universal positive constants B_7 *and* B_8 *and* $N_1 \in \mathbb{N}$ *such that for any* $N \geq N_1$, *with* $\mathbb{P}$ *probability at least* $1-\epsilon$,

$$\pi_{\exp(-\widehat{\gamma}\frac{r}{2})}(R) \leq \inf_\Theta R + B_7 c^{-\frac{1}{2\kappa-1}} \left[1 + \frac{B_8}{d} \log\left(\frac{\log(N)}{\epsilon} \right) \right]^{\frac{2\kappa}{2\kappa-1}} \left(\frac{d}{N} \right)^{\frac{\kappa}{2\kappa-1}},$$

where $\widehat{\gamma} \in \arg\max_{\gamma \in \Lambda_2} \max\bigl\{\beta \in \Lambda_2; B(\pi_{\exp(-\gamma \frac{r}{2})}, \beta, \gamma) \leq 0\bigr\}$, Λ_2 *being defined by equation (2.10, page 65) and* B *by Theorem 2.1.3 (page 54).*

We find the same rate of convergence as in Corollary 1.4.7 (page 40), but this time, we were able to provide an empirical posterior distribution $\pi_{\exp(-\widehat{\gamma}\frac{r}{2})}$ which achieves this rate adaptively in all the parameters (meaning in particular that we do not need to know d, c or κ). Moreover, as already mentioned, the power of N in this rate of convergence is known to be optimal in the worst case (see Mammen et al. (1999); Tsybakov (2004); Tsybakov et al. (2005), and more specifically in Audibert (2004b) — downloadable from its author's web page — Theorem 3.3, page 132).

2.1.5. Estimating the divergence of a posterior with respect to a Gibbs prior. Another interesting question is to estimate $\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr]$ using relative deviation inequalities. We follow here an idea to be found first in (Audibert, 2004b, page 93). Indeed, combining equation (2.3, page 52) with equation (2.1, page 51), we see that for any positive real parameters β and λ, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,

$$\begin{aligned}
\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] \leq \frac{\beta}{N\tanh(\frac{\gamma}{N})}\biggl\{\gamma\bigl[\rho(r) - \pi_{\exp(-\beta R)}(r)\bigr] \\
+ N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho\otimes\pi_{\exp(-\beta R)}(m') \\
+ \mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] - \log(\epsilon)\biggr\} + \mathcal{K}(\rho,\pi) - \mathcal{K}\bigl[\pi_{\exp(-\beta R)}, \pi\bigr] \\
\leq \mathcal{K}\bigl[\rho, \pi_{\exp[-\frac{\beta\gamma}{N\tanh(\frac{\gamma}{N})}r]}\bigr] + \frac{\beta}{N\tanh(\frac{\gamma}{N})}\Bigl\{\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] - \log(\epsilon)\Bigr\} \\
+ \log\biggl[\pi_{\exp[-\frac{\beta\gamma}{N\tanh(\frac{\gamma}{N})}r]}\Bigl\{\exp\Bigl[\frac{\beta}{\tanh(\frac{\gamma}{N})}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\Bigr]\Bigr\}\biggr].
\end{aligned}$$

We thus obtain

Theorem 2.1.18. *For any positive real constants* β *and* γ *such that* $\beta < N \times \tanh(\frac{\gamma}{N})$, *with* $\mathbb{P}$ *probability at least* $1-\epsilon$, *for any posterior distribution* $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,

$$\begin{aligned}
\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta R)}\bigr] \leq \left(1 - \frac{\beta}{N}\tanh\left(\frac{\gamma}{N}\right)^{-1}\right)^{-1} \\
\times \Biggl\{\mathcal{K}\bigl[\rho, \pi_{\exp[-\frac{\beta\gamma}{N}\tanh(\frac{\gamma}{N})^{-1}r]}\bigr] - \frac{\beta}{N\tanh(\frac{\gamma}{N})}\log(\epsilon) \\
+ \log\Bigl\{\pi_{\exp[-\frac{\beta\gamma}{N}\tanh(\frac{\gamma}{N})^{-1}r]}\Bigl[\exp\bigl\{\beta\tanh(\tfrac{\gamma}{N})^{-1}\log[\cosh(\tfrac{\gamma}{N})]\rho(m')\bigr\}\Bigr]\Bigr\}\Biggr\}.
\end{aligned}$$

This theorem provides another way of measuring over-fitting, since it gives an upper bound for $\mathcal{K}\bigl[\pi_{\exp[-\frac{\beta\gamma}{N}\tanh(\frac{\gamma}{N})^{-1}r]}, \pi_{\exp(-\beta R)}\bigr]$. It may be used in combination with Theorem 1.2.6 (page 11) as an alternative to Theorem 1.3.7 (page 21). It will also be used in the next section.

An alternative parametrization of the same result providing a simpler right-hand side is also useful:

COROLLARY 2.1.19. *For any positive real constants β and γ such that $\beta < \gamma$, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\rho : \Omega \to \mathcal{M}_+^1(\Theta)$,*

$$\mathcal{K}\bigl[\rho, \pi_{\exp[-N\frac{\beta}{\gamma}\tanh(\frac{\gamma}{N})R]}\bigr] \leq \left(1-\frac{\beta}{\gamma}\right)^{-1}\Biggl\{\mathcal{K}\bigl[\rho, \pi_{\exp(-\beta r)}\bigr] - \frac{\beta}{\gamma}\log(\epsilon) \\ + \log\Bigl\{\pi_{\exp(-\beta r)}\Bigl[\exp\bigl\{N\tfrac{\beta}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\}\Biggr\}.$$

2.2. Playing with two posterior and two local prior distributions

2.2.1. COMPARING TWO POSTERIOR DISTRIBUTIONS. Estimating the effective temperature of an estimator provides an efficient way to tune parameters in a model with parametric behaviour. On the other hand, it will not be fitted to choose between different models, especially when they are nested, because as we already saw in the case when Θ is a union of nested models, the prior distribution $\pi_{\exp(-\beta R)}$ does not provide an efficient localization of the parameter in this case, in the sense that $\pi_{\exp(-\beta R)}(R)$ does not go down to $\inf_\Theta R$ at the desired rate when β goes to $+\infty$, requiring a resort to partial localization.

Once some estimator (in the form of a posterior distribution) has been chosen in each sub-model, these estimators can be compared between themselves with the help of the relative bounds that we will establish in this section. It is also possible to choose several estimators in each sub-model, to tune parameters in the same time (like the inverse temperature parameter if we decide to use Gibbs posterior distributions in each sub-model).

From equation (2.2 page 52) (slightly modified by replacing $\pi \otimes \pi$ with $\pi^1 \otimes \pi^2$), we easily obtain

THEOREM 2.2.1. *For any positive real constant λ, for any prior distributions $\pi^1, \pi^2 \in \mathcal{M}_+^1(\Theta)$, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distributions ρ_1 and $\rho_2 : \Omega \to \mathcal{M}_+^1(\Theta)$,*

$$-N\log\Bigl\{1-\tanh\bigl(\tfrac{\lambda}{N}\bigr)\Bigl[\rho_2(R)-\rho_1(R)\Bigr]\Bigr\} \leq \lambda\bigl[\rho_2(r)-\rho_1(r)\bigr] \\ + N\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\rho_1\otimes\rho_2(m') \\ + \mathcal{K}\bigl(\rho_1,\pi^1\bigr) + \mathcal{K}\bigl(\rho_2,\pi^2\bigr) - \log(\epsilon).$$

This is where the entropy bound of the previous section enters into the game, providing a localized version of Theorem 2.2.1 (page 68). We will use the notation

$$\Xi_a(q) = \tanh(a)^{-1}\bigl[1-\exp(-aq)\bigr] \leq \frac{a}{\tanh(a)}q, \qquad a, q \in \mathbb{R}. \tag{2.11}$$

THEOREM 2.2.2. *For any $\epsilon \in)0,1($, any sequence of prior distributions $(\pi^i)_{i\in\mathbb{N}} \in \mathcal{M}_+^1(\Theta)^{\mathbb{N}}$, any probability distribution μ on $\mathbb{N}$, any atomic probability distribution ν on $\mathbb{R}_+$, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distributions $\rho_1, \rho_2 : \Omega \to \mathcal{M}_+^1(\Theta)$,*

$$\rho_2(R) - \rho_1(R) \leq B(\rho_1, \rho_2), \text{ where}$$

$$B(\rho_1,\rho_2) = \inf_{\lambda,\beta_1<\gamma_1,\beta_2<\gamma_2\in\mathbb{R}_+,i,j\in\mathbb{N}} \Xi_{\frac{\lambda}{N}}\Biggl\{ \bigl[\rho_2(r)-\rho_1(r)\bigr] + \tfrac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho_1\otimes\rho_2(m')$$
$$+ \frac{1}{\lambda\Bigl(1-\frac{\beta_1}{\gamma_1}\Bigr)}\biggl\{\mathcal{K}\bigl[\rho_1,\pi^i_{\exp(-\beta_1 r)}\bigr] + \log\Bigl\{\pi^i_{\exp(-\beta_1 r)}\Bigl[\exp\bigl\{\beta_1\tfrac{N}{\gamma_1}\log\bigl[\cosh(\tfrac{\gamma_1}{N})\bigr]\rho_1(m')\bigr\}\Bigr]\Bigr\} - \frac{\beta_1}{\gamma_1}\log\bigl[\nu(\gamma_1)\bigr]\biggr\}$$
$$+ \frac{1}{\lambda\Bigl(1-\frac{\beta_2}{\gamma_2}\Bigr)}\biggl\{\mathcal{K}\bigl[\rho_2,\pi^j_{\exp(-\beta_2 r)}\bigr] + \log\Bigl\{\pi^j_{\exp(-\beta_2 r)}\Bigl[\exp\bigl\{\beta_2\tfrac{N}{\gamma_2}\log\bigl[\cosh(\tfrac{\gamma_2}{N})\bigr]\rho_2(m')\bigr\}\Bigr]\Bigr\} - \frac{\beta_2}{\gamma_2}\log\bigl[\nu(\gamma_2)\bigr]\biggr\}$$
$$- \Bigl[\bigl(\tfrac{\gamma_1}{\beta_1}-1\bigr)^{-1} + \bigl(\tfrac{\gamma_2}{\beta_2}-1\bigr)^{-1} + 1\Bigr]\frac{\log\bigl[3^{-1}\nu(\beta_1)\nu(\beta_2)\nu(\lambda)\mu(i)\mu(j)\epsilon\bigr]}{\lambda}\Biggr\}.$$

The sequence of prior distributions $(\pi^i)_{i\in\mathbb{N}}$ should be understood to be typically supported by subsets of Θ corresponding to parametric sub-models, that is sub-models for which it is reasonable to expect that

$$\lim_{\beta\to+\infty}\beta\bigl[\pi^i_{\exp(-\beta R)}(R) - \operatorname{ess\,inf}_{\pi^i} R\bigr]$$

exists and is positive and finite. As there is no reason why the bound $B(\rho_1,\rho_2)$ provided by the previous theorem should be sub-additive (in the sense that $B(\rho_1,\rho_3)\leq B(\rho_1,\rho_2)+B(\rho_2,\rho_3)$), it is adequate to consider some workable subset $\mathcal{P}$ of posterior distributions (for instance the distributions of the form $\pi^i_{\exp(-\beta r)}$, $i\in\mathbb{N}$, $\beta\in\mathbb{R}_+$), and to define the sub-additive chained bound

$$\widetilde{B}(\rho,\rho') = \inf\Biggl\{\sum_{k=0}^{n-1} B(\rho_k,\rho_{k+1});\ n\in\mathbb{N}^*, (\rho_k)_{k=0}^n\in\mathcal{P}^{n+1}, \rho_0=\rho, \rho_n=\rho'\Biggr\},\quad \rho,\rho'\in\mathcal{P}. \tag{2.12}$$

Proposition 2.2.3. *With $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distributions $\rho_1,\rho_2\in\mathcal{P}$, $\rho_2(R)-\rho_1(R)\leq\widetilde{B}(\rho_1,\rho_2)$. Moreover for any posterior distribution $\rho_1\in\mathcal{P}$, any posterior distribution $\rho_2\in\mathcal{P}$ such that $\widetilde{B}(\rho_1,\rho_2)=\inf_{\rho_3\in\mathcal{P}}\widetilde{B}(\rho_1,\rho_3)$ is unimprovable with the help of $\widetilde{B}$ in $\mathcal{P}$ in the sense that $\inf_{\rho_3\in\mathcal{P}}\widetilde{B}(\rho_2,\rho_3)\geq 0$.*

Proof. The first assertion is a direct consequence of the previous theorem, so only the second assertion requires a proof: for any $\rho_3\in\mathcal{P}$, we deduce from the optimality of ρ_2 and the sub-additivity of $\widetilde{B}$ that

$$\widetilde{B}(\rho_1,\rho_2)\leq\widetilde{B}(\rho_1,\rho_3)\leq\widetilde{B}(\rho_1,\rho_2)+\widetilde{B}(\rho_2,\rho_3).$$

□

This proposition provides a way to improve a posterior distribution $\rho_1 \in \mathcal{P}$ by choosing $\rho_2 \in \arg\min_{\rho \in \mathcal{P}} \widetilde{B}(\rho_1, \rho)$ whenever $\widetilde{B}(\rho_1, \rho_2) < 0$. This improvement is proved by Proposition 2.2.3 to be one-step: the obtained improved posterior ρ_2 cannot be improved again using the same technique.

Let us give some examples of possible starting distributions ρ_1 for this improvement scheme: ρ_1 may be chosen as the best posterior Gibbs distribution according to Proposition 2.1.5 (page 56). More precisely, we may build from the prior distributions π^i, $i \in \mathbb{N}$, a global prior $\pi = \sum_{i \in \mathbb{N}} \mu(i) \pi^i$. We can then define the estimator of the inverse effective temperature as in Proposition 2.1.5 (page 56) and choose $\rho_1 \in \arg\min_{\rho \in \mathcal{P}} \widehat{\beta}(\rho)$, where $\mathcal{P}$ is as suggested above the set of posterior distributions

$$\mathcal{P} = \Bigl\{ \pi^i_{\exp(-\beta r)} ; \, i \in \mathbb{N}, \beta \in \mathbb{R}_+ \Bigr\}.$$

This starting point ρ_1 should already be pretty good, at least in an asymptotic perspective, the only gain in the rate of convergence to be expected bearing on spurious $\log(N)$ factors.

2.2.2. Elaborate uses of relative bounds between posteriors. More elaborate uses of relative bounds are described in the third section of the second chapter of Audibert (2004b), where an algorithm is proposed and analysed, which allows one to use relative bounds between two posterior distributions as a stand-alone estimation tool.

Let us give here some alternative way to address this issue. We will assume for simplicity and without great loss of generality that the working set of posterior distributions $\mathcal{P}$ is finite (so that among other things any ordering of it has a first element).

It is natural to define the estimated complexity of any given posterior distribution $\rho \in \mathcal{P}$ in our working set as the bound for $\inf_{i \in \mathbb{N}} \mathcal{K}(\rho, \pi^i)$ used in Theorem 2.2.1 (page 68). This leads to set (given some confidence level $1 - \epsilon$)

$$\begin{aligned} \mathcal{C}(\rho) = \inf_{\beta < \gamma \in \mathbb{R}_+, i \in \mathbb{N}} \biggl(1 - \frac{\beta}{\gamma} \biggr)^{-1} \biggl\{ & \mathcal{K}\bigl[\rho, \pi^i_{\exp(-\beta r)}\bigr] \\ & + \log \Bigl\{ \pi^i_{\exp(-\beta r)} \Bigl[\exp\bigl\{ \beta \tfrac{N}{\gamma} \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] \rho(m') \bigr\} \Bigr] \Bigr\} \\ & - \frac{\beta}{\gamma} \log\bigl[3^{-1} \nu(\gamma) \nu(\beta) \mu(i) \epsilon \bigr] \biggr\}. \end{aligned}$$

Let us moreover call $\gamma(\rho)$, $\beta(\rho)$ and $i(\rho)$ the values achieving this infimum, or nearly achieving it, which requires a slight change of the definition of $\mathcal{C}(\rho)$ to take this modification into account. For the sake of simplicity, we can assume without substantial loss of generality that the supports of ν and μ are large but finite, and thus that the minimum is reached.

To understand how this notion of complexity comes into play, it may be interesting to keep in mind that for any posterior distributions ρ and ρ' we can write the bound in Theorem 2.2.2 (page 68) as

$$B(\rho, \rho') = \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[\rho'(r) - \rho(r) + S_\lambda(\rho, \rho') \bigr], \tag{2.13}$$

where

$$S_\lambda(\rho, \rho') = S_\lambda(\rho', \rho) \leq \frac{N}{\lambda} \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr] \rho \otimes \rho'(m') + \frac{\mathcal{C}(\rho) + \mathcal{C}(\rho')}{\lambda} - \frac{\log(3^{-1}\epsilon)}{\lambda}$$

$$- \frac{\log\{\nu[\beta(\rho)]\mu[i(\rho)]\}}{\lambda(1-\frac{\beta(\rho')}{\gamma(\rho')})} - \frac{\log\{\nu[\beta(\rho')]\mu[i(\rho')]\}}{\lambda(1-\frac{\beta(\rho)}{\gamma(\rho)})}$$
$$- \Bigl[\bigl(\tfrac{\gamma(\rho)}{\beta(\rho)}-1\bigr)^{-1} + \bigl(\tfrac{\gamma(\rho')}{\beta(\rho')}-1\bigr)^{-1} + 1\Bigr] \frac{\log[\nu(\lambda)]}{\lambda}.$$

(Let us recall that the function Ξ is defined by equation (2.11, page 68).) Thus for any ρ, ρ' such that $B(\rho', \rho) > 0$, we can deduce from the monotonicity of $\Xi_{\frac{\lambda}{N}}$ that

$$\rho'(r) - \rho(r) \leq \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho, \rho'),$$

proving that the left-hand side is small, and consequently that $B(\rho, \rho')$ and its chained counterpart defined by equation (2.12, page 69) are small:

$$\widetilde{B}(\rho, \rho') \leq B(\rho, \rho') \leq \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}}\bigl[2S_\lambda(\rho, \rho')\bigr].$$

It is also worth noticing that $B(\rho, \rho')$ and $\widetilde{B}(\rho, \rho')$ are upper bounded in terms of variance and complexity only.

The presence of the ratios $\frac{\gamma(\rho)}{\beta(\rho)}$ should not be obnoxious, since their values should be automatically tamed by the fact that $\beta(\rho)$ and $\gamma(\rho)$ should make the estimate of the complexity of ρ optimal.

As an alternative, it is possible to restrict to set of parameter values β and γ such that, for some fixed constant $\zeta > 1$, the ratio $\frac{\gamma}{\beta}$ is bounded away from 1 by the inequality $\frac{\gamma}{\beta} \geq \zeta$. This leads to an alternative definition of $\mathcal{C}(\rho)$:

$$\mathcal{C}(\rho) = \inf_{\gamma \geq \zeta\beta \in \mathbb{R}_+, i \in \mathbb{N}} \left(1 - \frac{\beta}{\gamma}\right)^{-1} \biggl\{ \mathcal{K}\bigl[\rho, \pi^i_{\exp(-\beta r)}\bigr]$$
$$+ \log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\bigl\{\beta\tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\}$$
$$- \frac{\beta}{\gamma}\log\bigl[3^{-1}\nu(\gamma)\nu(\beta)\mu(i)\epsilon\bigr]\biggr\} - \frac{\log[\nu(\beta)\mu(i)]}{(1-\zeta^{-1})} - \frac{\log(3^{-1}\epsilon)}{2}.$$

We can even push simplification a step further, postponing the optimization of the ratio $\frac{\gamma}{\beta}$, and setting it to the fixed value ζ. This leads us to adopt the definition

$$\mathcal{C}(\rho) = \inf_{\beta \in \mathbb{R}_+, i \in \mathbb{N}} \bigl(1 - \zeta^{-1}\bigr)^{-1} \biggl\{ \mathcal{K}\bigl[\rho, \pi^i_{\exp(-\beta r)}\bigr]$$
$$+ \log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\bigl\{\tfrac{N}{\zeta}\log\bigl[\cosh(\tfrac{\zeta\beta}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\}\biggr\}$$
$$- \frac{\zeta+1}{\zeta-1}\Bigl\{\log\bigl[\nu(\beta)\mu(i)\bigr] + 2^{-1}\log(3^{-1}\epsilon)\Bigr\}. \tag{2.14}$$

With either of these modified definitions of the complexity $\mathcal{C}(\rho)$, we get the upper bound

$$S_\lambda(\rho, \rho') \leq \widetilde{S}_\lambda(\rho, \rho') \overset{\text{def}}{=} \frac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho \otimes \rho'(m')$$
$$+ \frac{1}{\lambda}\Bigl\{\mathcal{C}(\rho) + \mathcal{C}(\rho') - \frac{\zeta+1}{\zeta-1}\log\bigl[\nu(\lambda)\bigr]\Bigr\}. \tag{2.15}$$

With these definitions, we have for any posterior distributions ρ and ρ'

$$B(\rho, \rho') \leq \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \Bigl\{ \rho'(r) - \rho(r) + \widetilde{S}_{\lambda}(\rho, \rho') \Bigr\}.$$

Consequently in the case when $B(\rho', \rho) > 0$, we get

$$\widetilde{B}(\rho, \rho') \leq B(\rho, \rho') \leq \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 \widetilde{S}_{\lambda}(\rho, \rho') \bigr].$$

To select some nearly optimal posterior distribution in $\mathcal{P}$, it is appropriate to order the posterior distributions of $\mathcal{P}$ according to increasing values of their complexity $\mathcal{C}(\rho)$ and consider some indexation $\mathcal{P} = \{\rho_1, \dots, \rho_M\}$, where $\mathcal{C}(\rho_k) \leq \mathcal{C}(\rho_{k+1})$, $1 \leq k < M$.

Let us now consider for each $\rho_k \in \mathcal{P}$ the first posterior distribution in $\mathcal{P}$ which cannot be proved to be worse than ρ_k according to the bound $\widetilde{B}$:

$$t(k) = \min \Bigl\{ j \in \{1, \dots M\} \,:\, \widetilde{B}(\rho_j, \rho_k) > 0 \Bigr\}. \tag{2.16}$$

In this definition, which uses the chained bound defined by equation (2.12, page 69), it is appropriate to assume by convention that $\widetilde{B}(\rho, \rho) = 0$, for any posterior distribution ρ. Let us now define our estimated best $\rho \in \mathcal{P}$ as $\rho_{\widehat{k}}$, where

$$\widehat{k} = \min(\arg\max t). \tag{2.17}$$

Thus we take the posterior with smallest complexity which can be proved to be better than the largest starting interval of $\mathcal{P}$ in terms of estimated relative classification error.

The following theorem is a simple consequence of the chosen optimisation scheme. It is valid for any arbitrary choice of the complexity function $\rho \mapsto \mathcal{C}(\rho)$.

THEOREM 2.2.4. *Let us put $\widehat{t} = t(\widehat{k})$, where t is defined by equation (2.16) and $\widehat{k}$ is defined by equation (2.17). With $\mathbb{P}$ probability at least $1 - \epsilon$,*

$$\rho_{\widehat{k}}(R) \leq \rho_j(R) + \begin{cases} 0, & 1 \leq j < \widehat{t}, \\ \widetilde{B}(\rho_j, \rho_{t(j)}), & \widehat{t} \leq j < \widehat{k}, \\ \widetilde{B}(\rho_j, \rho_{\widehat{t}}) + \widetilde{B}(\rho_{\widehat{t}}, \rho_{\widehat{k}}), & j \in (\arg\max t), \\ \widetilde{B}(\rho_j, \rho_{\widehat{k}}), & j \in \bigl\{\widehat{k}+1, \dots, M\bigr\} \setminus (\arg\max t), \end{cases}$$

where the chained bound $\widetilde{B}$ is defined from the bound of Theorem 2.2.2 (page 68) by equation (2.12, page 69). In the mean time, for any j such that $\widehat{t} \leq j < \widehat{k}$, $t(j) < \widehat{t} = \max t$, because $j \notin (\arg\max t)$. Thus

$$\rho_{\widehat{k}}(R) \leq \rho_{t(j)}(R) \leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 S_{\lambda}(\rho_j, \rho_{t(j)}) \bigr]$$

$$\textit{while } \rho_{t(j)}(r) \leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_{\lambda}(\rho_j, \rho_{t(j)}),$$

where the function Ξ is defined by equation (2.11, page 68) and S_λ is defined by equation (2.13, page 70). For any $j \in (\arg\max t)$, (including notably $\widehat{k}$),

$$B(\rho_{\widehat{t}}, \rho_j) \geq \widetilde{B}(\rho_{\widehat{t}}, \rho_j) > 0,$$

$$B(\rho_j, \rho_{\widehat{t}}) \geq \widetilde{B}(\rho_j, \rho_{\widehat{t}}) > 0,$$

so in this case

$$\begin{aligned} \rho_{\widehat{k}}(R) &\leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \Bigl[S_\lambda(\rho_j, \rho_{\widehat{t}}) + S_\lambda(\rho_{\widehat{t}}, \rho_{\widehat{k}}) + S_\lambda(\rho_j, \rho_{\widehat{k}}) \Bigr], \\ \text{while } \rho_{\widehat{t}}(r) &\leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{\widehat{t}}), \\ \rho_{\widehat{k}}(r) &\leq \rho_{\widehat{t}}(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_{\widehat{t}}, \rho_{\widehat{k}}), \\ \text{and } \rho_{\widehat{t}}(R) &\leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 S_\lambda(\rho_j, \rho_{\widehat{t}}) \bigr]. \end{aligned}$$

Finally in the case when $j \in \bigl\{ \widehat{k} + 1, \dots, M \bigr\} \setminus (\arg\max t)$*, due to the fact that in particular* $j \notin (\arg\max t)$*,*

$$B(\rho_{\widehat{k}}, \rho_j) \geq \widetilde{B}(\rho_{\widehat{k}}, \rho_j) > 0.$$

Thus in this last case

$$\begin{aligned} \rho_{\widehat{k}}(R) &\leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 S_\lambda(\rho_j, \rho_{\widehat{k}}) \bigr], \\ \text{while } \rho_{\widehat{k}}(r) &\leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{\widehat{k}}). \end{aligned}$$

Thus for any $j = 1, \dots, M$*,* $\rho_{\widehat{k}}(R) - \rho_j(R)$ *is bounded from above by an empirical quantity involving only variance and entropy terms of posterior distributions* ρ_ℓ *such that* $\ell \leq j$*, and therefore such that* $\mathcal{C}(\rho_\ell) \leq \mathcal{C}(\rho_j)$*. Moreover, these distributions* ρ_ℓ *are such that* $\rho_\ell(r) - \rho_j(r)$ *and* $\rho_\ell(R) - \rho_j(R)$ *have an empirical upper bound of the same order as the bound stated for* $\rho_{\widehat{k}}(R) - \rho_j(R)$ *— namely the bound for* $\rho_\ell(r) - \rho_j(r)$ *is in all circumstances not greater than* $\Xi_{\frac{\lambda}{N}}^{-1}$ *applied to the bound stated for* $\rho_{\widehat{k}}(R) - \rho_j(R)$*, whereas the bound for* $\rho_\ell(R) - \rho_j(R)$ *is always smaller than two times the bound stated for* $\rho_{\widehat{k}}(R) - \rho_j(R)$*. This shows that variance terms are between posterior distributions whose empirical as well as expected error rates cannot be much larger than those of* ρ_j*.*

Let us remark that the estimation scheme described in this theorem is very general, the same method can be used as soon as some *confidence interval* for the relative expected risks

$$-B(\rho_2, \rho_1) \leq \rho_2(R) - \rho_1(R) \leq B(\rho_1, \rho_2) \text{ with } \mathbb{P} \text{ probability at least } 1 - \epsilon,$$

is available. The definition of the complexity is arbitrary, and could in an abstract context be chosen as

$$\mathcal{C}(\rho_1) = \inf_{\rho_2 \neq \rho_1} B(\rho_1, \rho_2) + B(\rho_2, \rho_1).$$

Proof. The case when $1 \leq j < \widehat{t}$ is straightforward from the definitions: when $j < \widehat{t}$, $\widetilde{B}(\rho_j, \rho_{\widehat{k}}) \leq 0$ and therefore $\rho_{\widehat{k}}(R) \leq \rho_j(R)$.

In the second case, that is when $\widehat{t} \leq j < \widehat{k}$, j cannot be in $\arg\max t$, because of the special choice of $\widehat{k}$ in $\arg\max t$. Thus $t(j) < \widehat{t}$ and we deduce from the first case that

$$\rho_{\widehat{k}}(R) \leq \rho_{t(j)}(R) \leq \rho_j(R) + \widetilde{B}(\rho_j, \rho_{t(j)}).$$

Moreover, we see from the defintion of t that $\widetilde{B}(\rho_{t(j)}, \rho_j) > 0$, implying

$$\rho_{t(j)}(r) \leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{t(j)}),$$

and therefore that

$$\rho_{\widehat{k}}(R) \leq \rho_j(R) + \inf_{\lambda} \Xi_{\frac{\lambda}{N}} \bigl[2 S_\lambda(\rho_j, \rho_{t(j)})\bigr].$$

In the third case j belongs to $\arg\max t$. In this case, we are not sure that $\widetilde{B}(\rho_{\widehat{k}}, \rho_j) > 0$, and it is appropriate to involve $\widehat{t}$, which is the index of the first posterior distribution which cannot be improved by $\rho_{\widehat{k}}$, implying notably that $\widetilde{B}(\rho_{\widehat{t}}, \rho_k) > 0$ for any $k \in \arg\max t$. On the other hand, $\rho_{\widehat{t}}$ cannot either improve any posterior distribution ρ_k with $k \in (\arg\max t)$, because this would imply for any $\ell < \widehat{t}$ that $\widetilde{B}(\rho_\ell, \rho_{\widehat{t}}) \leq \widetilde{B}(\rho_\ell, \rho_k) + \widetilde{B}(\rho_k, \rho_{\widehat{t}}) \leq 0$, and therefore that $t(\widehat{t}) \geq \widehat{t} + 1$, in contradiction of the fact that $\widehat{t} = \max t$. Thus $\widetilde{B}(\rho_k, \rho_{\widehat{t}}) > 0$, and these two remarks imply that

$$\begin{aligned}
\rho_{\widehat{t}}(r) &\leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{\widehat{t}}),\\
\rho_{\widehat{k}}(r) &\leq \rho_{\widehat{t}}(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_{\widehat{t}}, \rho_{\widehat{k}})\\
&\leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{\widehat{t}}) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_{\widehat{t}}, \rho_{\widehat{k}}),
\end{aligned}$$

and consequently also that

$$\begin{aligned}
\rho_{\widehat{k}}(R) &\leq \rho_j(R) + \widetilde{B}(\rho_j, \rho_{\widehat{k}})\\
&\leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \Bigl[S_\lambda(\rho_j, \rho_{\widehat{t}}) + S_\lambda(\rho_{\widehat{t}}, \rho_{\widehat{k}}) + S_\lambda(\rho_j, \rho_{\widehat{k}}) \Bigr]
\end{aligned}$$

and that

$$\rho_{\widehat{t}}(R) \leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 S_\lambda(\rho_j, \rho_{\widehat{t}})\bigr] \leq \rho_j(R) + 2 \inf_{\lambda \in \mathbb{R}_+} 2 \Xi_{\frac{\lambda}{N}} \bigl[S_\lambda(\rho_j, \rho_{\widehat{t}})\bigr],$$

the last inequality being due to the fact that $\Xi_{\frac{\lambda}{N}}$ is a concave function. Let us notice that it may be the case that $\widehat{k} < \widehat{t}$, but that only the case when $j \geq \widehat{t}$ is to be considered, since otherwise we already know that $\rho_{\widehat{k}}(R) \leq \rho_j(R)$.

In the fourth case, j is greater than $\widehat{k}$, and the complexity of ρ_j is larger than the complexity of $\rho_{\widehat{k}}$. Moreover, j is not in $\arg\max t$, and thus $\widetilde{B}(\rho_{\widehat{k}}, \rho_j) > 0$, because otherwise, the sub-additivity of $\widetilde{B}$ would imply that $\widetilde{B}(\rho_\ell, \rho_j) \leq 0$ for any $\ell \leq \widehat{t}$ and therefore that $t(j) \geq \widehat{t} = \max t$. Therefore

$$\rho_{\widehat{k}}(r) \leq \rho_j(r) + \inf_{\lambda \in \mathbb{R}_+} S_\lambda(\rho_j, \rho_{\widehat{k}}),$$

and

$$\rho_{\widehat{k}}(R) \leq \rho_j(R) + \widetilde{B}(\rho_j, \rho_{\widehat{k}}) \leq \rho_j(R) + \inf_{\lambda \in \mathbb{R}_+} \Xi_{\frac{\lambda}{N}} \bigl[2 S_\lambda(\rho_j, \rho_{\widehat{k}})\bigr].$$

□

2.2.3. Analysis of relative bounds. Let us start our investigation of the theoretical properties of the algorithm described in Theorem 2.2.4 (page 72) by computing some non-random upper bounds for $B(\rho,\rho')$, the bound of Theorem 2.2.2 (page 68), and $\mathcal{C}(\rho)$, the complexity factor defined by equation (2.14, page 71), for any $\rho,\rho'\in\mathcal{P}$.

This analysis will be done in the case when

$$\mathcal{P}=\Bigl\{\pi^i_{\exp(-\beta r)}\,:\,\nu(\beta)>0,\mu(i)>0\Bigr\},$$

in which it will be possible to get some control on the randomness of any $\rho\in\mathcal{P}$, in addition to controlling the other random expressions appearing in the definition of $B(\rho,\rho')$, $\rho,\rho'\in\mathcal{P}$. We will also use a simpler choice of complexity function, removing from equation (2.14 page 71) the optimization in i and β and using instead the definition

$$\mathcal{C}(\pi^i_{\exp(-\beta r)})\overset{\text{def}}{=}\bigl(1-\zeta^{-1}\bigr)^{-1}\log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\Bigl\{\tfrac{N}{\zeta}\log\bigl[\cosh\bigl(\tfrac{\zeta\beta}{N}\bigr)\bigr]\pi^i_{\exp(-\beta r)}(m')\Bigr\}\Bigr]\Bigr\}+\frac{\zeta+1}{\zeta-1}\log\bigl[\nu(\beta)\mu(i)\bigr]. \tag{2.18}$$

With this definition,

$$S_\lambda(\pi^i_{\exp(-\beta r)},\pi^j_{\exp(-\beta' r)})\leq\frac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\pi^i_{\exp(-\beta r)}\otimes\pi^j_{\exp(-\beta' r)}(m')+\frac{\mathcal{C}\bigl[\pi^i_{\exp(-\beta r)}\bigr]+\mathcal{C}\bigl[\pi^j_{\exp(-\beta' r)}\bigr]}{\lambda}+\frac{(\zeta+1)}{(\zeta-1)\lambda}\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr],$$

where S_λ is defined by equation (2.13, page 70), so that

$$B\bigl[\pi^i_{\exp(-\beta r)},\pi^j_{\exp(-\beta' r)}\bigr]=\inf_{\lambda\in\mathbb{R}_+}\Xi_{\frac{\lambda}{N}}\Bigl\{\pi^j_{\exp(-\beta' r)}(r)-\pi^i_{\exp(-\beta r)}(r)+S_\lambda\bigl[\pi^i_{\exp(-\beta r)},\pi^j_{\exp(-\beta r)}\bigr]\Bigr\}.$$

Let us successively bound the various random factors entering into the definition of $B\bigl[\pi^i_{\exp(-\beta r)},\pi^j_{\exp(-\beta' r)}\bigr]$. The quantity $\pi^j_{\exp(-\beta' r)}(r)-\pi^i_{\exp(-\beta r)}(r)$ can be bounded using a slight adaptation of Proposition 2.1.11 (page 58).

Proposition 2.2.5. *For any positive real constants λ,λ' and γ, with $\mathbb{P}$ probability at least $1-\eta$, for any positive real constants β, β' such that $\beta<\lambda\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$ and $\beta'>\lambda'\frac{\gamma}{N}\sinh(\frac{\gamma}{N})^{-1}$,*

$$\pi^j_{\exp(-\beta' r)}(r)-\pi^i_{\exp(-\beta r)}(r)\leq\pi^j_{\exp(-\lambda' R)}\otimes\pi^i_{\exp(-\lambda R)}\bigl[\Psi_{-\frac{\gamma}{N}}(R',M')\bigr]+\frac{\log(\frac{3}{\eta})}{\gamma}+\frac{C^j(\lambda',\gamma)+\log(\frac{3}{\eta})}{\frac{N\beta'}{\lambda'}\sinh(\frac{\gamma}{N})-\gamma}+\frac{C^i(\lambda,\gamma)+\log(\frac{3}{\eta})}{\gamma-\frac{N\beta}{\lambda}\sinh(\frac{\gamma}{N})},$$

where

$$C^i(\lambda,\gamma) \overset{\text{def}}{=} \log\biggl\{ \int_\Theta \exp\biggl[-\gamma \int_\Theta \Bigl\{ \Psi_{\frac{\gamma}{N}}\bigl[R'(\theta_1,\theta_2), M'(\theta_1,\theta_2)\bigr] \\ - \tfrac{N}{\gamma}\sinh(\tfrac{\gamma}{N}) R'(\theta_1,\theta_2)\Bigr\} \pi^i_{\exp(-\lambda R)}(d\theta_2)\biggr] \pi^i_{\exp(-\lambda R)}(d\theta_1)\biggr\} \\ \leq \log\biggl\{ \pi^i_{\exp(-\lambda R)}\Bigl[\exp\Bigl\{2N\sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \pi^i_{\exp(-\lambda R)}\bigl(M'\bigr)\Bigr\}\Bigr]\biggr\}.$$

As for $\pi^i_{\exp(-\beta r)} \otimes \pi^j_{\exp(-\beta' r)}(m')$, we can write with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distributions ρ and $\rho' : \Omega \to \mathcal{M}^1_+(\Theta)$,

$$\gamma \rho \otimes \rho'(m') \leq \log\Bigl[\pi^i_{\exp(-\lambda R)} \otimes \pi^j_{\exp(-\lambda' R)}\bigl\{\exp\bigl[\gamma \Phi_{-\frac{\gamma}{N}}(M')\bigr]\bigr\}\Bigr] \\ + \mathcal{K}\bigl[\rho, \pi^i_{\exp(-\lambda R)}\bigr] + \mathcal{K}\bigl[\rho', \pi^j_{\exp(-\lambda' R)}\bigr] - \log(\eta).$$

We can then replace λ with $\beta \frac{N}{\lambda}\sinh(\frac{\lambda}{N})$ and use Theorem 2.1.12 (page 60) to get

Proposition 2.2.6. *For any positive real constants γ, λ, λ', β and β', with $\mathbb{P}$ probability $1-\eta$,*

$$\gamma\rho\otimes\rho'(m') \\ \leq \log\Bigl[\pi^i_{\exp[-\beta\frac{N}{\lambda}\sinh(\frac{\lambda}{N})R]} \otimes \pi^j_{\exp[-\beta'\frac{N}{\lambda'}\sinh(\frac{\lambda'}{N})R]}\bigl\{\exp\bigl[\gamma\Phi_{-\frac{\gamma}{N}}(M')\bigr]\bigr\}\Bigr] \\ + \frac{\mathcal{K}\bigl[\rho,\pi^i_{\exp(-\beta r)}\bigr]}{1-\frac{\beta}{\lambda}} + \frac{C^i\bigl[\beta\frac{N}{\lambda}\sinh(\frac{\lambda}{N}),\lambda\bigr] - \log(\frac{\eta}{3})}{\frac{\lambda}{\beta}-1} \\ + \frac{\mathcal{K}\bigl[\rho',\pi^j_{\exp(-\beta' r)}\bigr]}{1-\frac{\beta'}{\lambda'}} + \frac{C^j\bigl[\beta'\frac{N}{\lambda'}\sinh(\frac{\lambda'}{N}),\lambda'\bigr] - \log(\frac{\eta}{3})}{\frac{\lambda}{\beta'}-1} - \log(\tfrac{\eta}{3}).$$

The last random factor in $B(\rho,\rho')$ that we need to upper bound is

$$\log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\bigl\{\beta\tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi^i_{\exp(-\beta r)}(m')\bigr\}\Bigr]\Bigr\}.$$

A slight adaptation of Proposition 2.1.13 (page 60) shows that with $\mathbb{P}$ probability at least $1-\eta$,

$$\log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\bigl\{\beta\tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi^i_{\exp(-\beta r)}(m')\bigr\}\Bigr]\Bigr\} \\ \leq \frac{2\beta}{\gamma} C^i\bigl[\tfrac{N\beta}{\gamma}\sinh(\tfrac{\gamma}{N}),\gamma\bigr] + \bigl(1-\tfrac{\beta}{\gamma}\bigr)\log\biggl\{\Bigl(\pi^i_{\exp[-\frac{N\beta}{\gamma}\sinh(\frac{\gamma}{N})R]}\Bigr)^{\otimes 2}\biggl[\\ \exp\biggl(\frac{N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]}{\frac{\gamma}{\beta}-1}\Phi_{-\frac{\log[\cosh(\frac{\gamma}{N})]}{\frac{\gamma}{\beta}-1}} \circ M'\biggr)\biggr]\biggr\} \\ + \bigl(1+\tfrac{\beta}{\gamma}\bigr)\log(\tfrac{2}{\eta}),$$

where as usual Φ is the function defined by equation (1.1, page 2). This leads us to define for any $i, j \in \mathbb{N}$, any $\beta, \beta' \in \mathbb{R}_+$,

$$(2.19)\quad \overline{\mathcal{C}}(i,\beta) \overset{\text{def}}{=} \frac{2}{\zeta-1} C^i\Big[\tfrac{N}{\zeta}\sinh(\tfrac{\zeta\beta}{N}), \zeta\beta\Big] \\ + \log\Bigg\{ \big(\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\big)^{\otimes 2}\Bigg[\\ \exp\Bigg(\frac{N\log\big[\cosh(\frac{\zeta\beta}{N})\big]}{\zeta-1}\Phi_{-\frac{\log[\cosh(\frac{\zeta\beta}{N})]}{\zeta-1}}\circ M'\Bigg)\Bigg]\Bigg\} \\ - \frac{\zeta+1}{\zeta-1}\Big\{2\log\big[\nu(\beta)\mu(i)\big] + \log\big(\tfrac{\eta}{2}\big)\Big\}.$$

Recall that the definition of $C^i(\lambda,\gamma)$ is to be found in Proposition 2.2.5, page 75. Let us remark that, since

$$\exp\big[Na\Phi_{-a}(p)\big] = \exp\Big\{N\log\Big[1+\big[\exp(a)-1\big]p\Big]\Big\} \\ \leq \exp\Big\{N\big[\exp(a)-1\big]p\Big\}, \quad p\in(0,1), a\in\mathbb{R},$$

we have

$$\overline{\mathcal{C}}(i,\beta) \leq \frac{2}{\zeta-1}\log\Bigg\{\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\Bigg[\\ \exp\Big\{2N\sinh\big(\tfrac{\zeta\beta}{2N}\big)^2\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\big(M'\big)\Big\}\Bigg]\Bigg\} \\ + \log\Bigg\{\Big(\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\Big)^{\otimes 2}\Bigg[\\ \exp\Big\{N\Big[\exp\{(\zeta-1)^{-1}\log\big[\cosh\big(\tfrac{\zeta\beta}{N}\big)\big]\}-1\Big]M'\Big\}\Bigg]\Bigg\} \\ - \frac{\zeta+1}{\zeta-1}\Big\{2\log\big[\nu(\beta)\mu(i)\big]+\log\big(\tfrac{\eta}{2}\big)\Big\}.$$

Let us put

$$\overline{S}_\lambda\big[(i,\beta),(j,\beta')\big] \overset{\text{def}}{=} \frac{N}{\lambda}\log\big[\cosh(\tfrac{\lambda}{N})\big]\inf_{\gamma\in\mathbb{R}_+}\gamma^{-1}\Bigg\{ \\ \log\Bigg[\Big(\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\otimes\pi^j_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N})R]}\Big)\Big\{\exp\big[\gamma\Phi_{-\frac{\gamma}{N}}(M')\big]\Big\}\Bigg] \\ + \frac{C^i\big[\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N}),\zeta\beta\big]-\log(\frac{\overline{\eta}}{3})}{\zeta-1} \\ + \frac{C^j\big[\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N}),\zeta\beta'\big]-\log(\frac{\overline{\eta}}{3})}{\zeta-1} - \log(\tfrac{\overline{\eta}}{3})\Bigg\} \\ + \frac{1}{\lambda}\Bigg[\overline{\mathcal{C}}(i,\beta)+\overline{\mathcal{C}}(j,\beta')-\frac{\zeta+1}{\zeta-1}\log\big[3^{-1}\nu(\lambda)\epsilon\big]\Bigg],$$

where

$$\overline{\eta} = \nu(\gamma)\nu(\beta)\nu(\beta')\mu(i)\mu(j)\eta.$$

Let us remark that

$$\overline{S}_\lambda\bigl[(i,\beta),(j,\beta')\bigr] \leq \inf_{\gamma\in\mathbb{R}_+} \frac{\lambda}{2N\gamma}\log\biggl[\Bigl(\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\otimes\pi^j_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N})R]}\Bigr)\Bigl\{ \exp\Bigl[N\bigl[\exp\bigl(\tfrac{\gamma}{N}\bigr)-1\bigr]M'\Bigr]\Bigr\}\biggr]$$
$$+\left(\frac{\lambda}{2N\gamma(\zeta-1)}+\frac{2}{\lambda(\zeta-1)}\right)\log\Bigl\{\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\Bigl[\exp\Bigl\{2N\sinh\bigl(\tfrac{\zeta\beta}{2N}\bigr)^2\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\bigl(M'\bigr)\Bigr\}\Bigr]\Bigr\}$$
$$+\lambda^{-1}\log\Bigl\{\Bigl(\pi^i_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta}{N})R]}\Bigr)^{\otimes 2}\Bigl[\exp\Bigl\{N\Bigl[\exp\bigl\{(\zeta-1)^{-1}\log\bigl[\cosh\bigl(\tfrac{\zeta\beta}{N}\bigr)\bigr]\bigr\}-1\Bigr]M'\Bigr\}\Bigr]\Bigr\}$$
$$+\left(\frac{\lambda}{2N\gamma(\zeta-1)}+\frac{2}{\lambda(\zeta-1)}\right)\log\Bigl\{\pi^j_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N})R]}\Bigl[\exp\Bigl\{2N\sinh\bigl(\tfrac{\zeta\beta'}{2N}\bigr)^2\pi^j_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N})R]}\bigl(M'\bigr)\Bigr\}\Bigr]\Bigr\}$$
$$+\lambda^{-1}\log\Bigl\{\Bigl(\pi^j_{\exp[-\frac{N}{\zeta}\sinh(\frac{\zeta\beta'}{N})R]}\Bigr)^{\otimes 2}\Bigl[\exp\Bigl\{N\Bigl[\exp\bigl\{(\zeta-1)^{-1}\log\bigl[\cosh\bigl(\tfrac{\zeta\beta'}{N}\bigr)\bigr]\bigr\}-1\Bigr]M'\Bigr\}\Bigr]\Bigr\}$$
$$-\frac{(\zeta+1)\lambda}{2N(\zeta-1)\gamma}\log\bigl[3^{-1}\nu(\gamma)\nu(\beta)\nu(\beta')\mu(i)\mu(j)\eta\bigr]$$
$$-\frac{(\zeta+1)}{(\zeta-1)\lambda}\Bigl(2\log\bigl[2^{-1}\nu(\beta)\nu(\beta')\mu(i)\mu(j)\eta\bigr]+\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr]\Bigr).$$

Let us define accordingly

$$\overline{B}\bigl[(i,\beta),(j,\beta')\bigr] \overset{\text{def}}{=} \inf_{\lambda}\Xi_{\frac{\lambda}{N}}\biggl\{\inf_{\alpha,\gamma,\alpha',\gamma'}\biggl[\pi^j_{\exp(-\alpha' R)}\otimes\pi^i_{\exp(-\alpha R)}\bigl[\Psi_{-\frac{\lambda}{N}}(R',M')\bigr]$$
$$-\frac{\log\bigl(\frac{\widetilde{\eta}}{3}\bigr)}{\lambda}+\frac{C^j(\alpha',\gamma')-\log\bigl(\frac{\widetilde{\eta}}{3}\bigr)}{\frac{N\beta'}{\alpha'}\sinh(\frac{\gamma'}{N})-\gamma'}+\frac{C^i(\alpha,\gamma)-\log\bigl(\frac{\widetilde{\eta}}{3}\bigr)}{\gamma-\frac{N\beta}{\alpha}\sinh(\frac{\gamma}{N})}\biggr]$$
$$+\overline{S}_\lambda\bigl[(i,\beta),(j,\beta')\bigr]\biggr\},$$

where

$$\widetilde{\eta}=\nu(\lambda)\nu(\alpha)\nu(\gamma)\nu(\beta)\nu(\alpha')\nu(\gamma')\nu(\beta')\mu(i)\mu(j)\eta.$$

Proposition 2.2.7.

- *With $\mathbb{P}$ probability at least $1-\eta$, for any $\beta\in\mathbb{R}_+$ and $i\in\mathbb{N}$,* $\mathcal{C}(\pi^i_{\exp(-\beta r)})\leq\overline{\mathcal{C}}(i,\beta)$;
- *With $\mathbb{P}$ probability at least $1-3\eta$, for any $\lambda,\beta,\beta'\in\mathbb{R}_+$, any $i,j\in\mathbb{N}$,* $S_\lambda\bigl[(i,\beta),(j,\beta')\bigr]\leq\overline{S}_\lambda\bigl[(i,\beta),(j,\beta')\bigr]$;

- *With* $\mathbb{P}$ *probability at least* $1 - 4\eta$, *for any* $i, j \in \mathbb{N}$, *any* $\beta, \beta' \in \mathbb{R}_+$, $B(\pi^i_{\exp(-\beta r)}, \pi^j_{\exp(-\beta' r)}) \leq \overline{B}\bigl[(i,\beta),(j,\beta')\bigr]$.

It is also interesting to find a non-random lower bound for $\mathcal{C}(\pi^i_{\exp(-\beta r)})$. Let us start from the fact that with $\mathbb{P}$ probability at least $1-\eta$,

$$\pi^i_{\exp(-\alpha R)} \otimes \pi^i_{\exp(-\alpha R)}\bigl[\Phi_{\frac{\gamma'}{N}}(M')\bigr] \\ \leq \pi^i_{\exp(-\alpha R)} \otimes \pi^i_{\exp(-\alpha R)}(m') - \frac{\log(\eta)}{\gamma'}.$$

On the other hand, we already proved that with $\mathbb{P}$ probability at least $1-\eta$,

$$0 \leq \left(1 - \frac{\alpha}{N \tanh(\frac{\lambda}{N})}\right)\mathcal{K}\bigl[\rho, \pi^i_{\exp(-\alpha R)}\bigr] \\ \leq \frac{\alpha}{N\tanh(\frac{\lambda}{N})}\biggl\{\lambda\bigl[\rho(r) - \pi^i_{\exp(\alpha R)}(r)\bigr] \\ + N\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho\otimes\pi^i_{\exp(-\alpha R)}(m') - \log(\eta)\biggr\} \\ + \mathcal{K}\bigl(\rho,\pi^i\bigr) - \mathcal{K}\bigl(\pi^i_{\exp(-\alpha R)},\pi^i\bigr).$$

Thus for any $\xi > 0$, putting $\beta = \frac{\alpha\lambda}{N\tanh(\frac{\lambda}{N})}$, with $\mathbb{P}$ probability at least $1-\eta$,

$$\xi\pi^i_{\exp(-\alpha R)} \otimes \pi^i_{\exp(-\alpha R)}\bigl[\Phi_{\frac{\gamma'}{N}}(M')\bigr] \\ \leq \pi^i_{\exp(-\alpha R)}\biggl\{\log\biggl[\pi^i_{\exp(-\beta r)}\biggl\{ \\ \exp\Bigl[\beta\tfrac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\pi^i_{\exp(-\beta r)}(m') + \xi m'\Bigr]\biggr\}\biggr]\biggr\} \\ - \left(\frac{\beta}{\lambda} + \frac{\xi}{\gamma'}\right)\log\left(\frac{\eta}{2}\right) \\ \leq \log\biggl\{\pi^i_{\exp(-\beta r)}\biggl[\exp\Bigl\{\beta\tfrac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\pi^i_{\exp(-\beta r)}(m')\Bigr\} \\ \times \pi^i_{\exp(-\beta r)}\Bigl\{\exp\Bigl[\beta\tfrac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\pi^i_{\exp(-\beta r)}(m') + \xi m'\Bigr]\Bigr\}\biggr]\biggr\} \\ - \left(2\frac{\beta}{\lambda} + \frac{\xi}{\gamma'}\right)\log\left(\frac{\eta}{2}\right) \\ \leq 2\log\biggl\{\pi^i_{\exp(-\beta r)}\biggl[\exp\Bigl\{\Bigl[\xi + \beta\tfrac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\Bigr]\pi^i_{\exp(-\beta r)}(m')\Bigr\}\biggr]\biggr\} \\ - \left(2\frac{\beta}{\lambda} + \frac{\xi}{\gamma'}\right)\log\left(\frac{\eta}{2}\right) \\ \leq 2\log\biggl\{\pi^i_{\exp(-\beta r)}\biggl[\exp\Bigl\{\Bigl[\xi + \tfrac{\beta\lambda}{2N}\Bigr]\pi^i_{\exp(-\beta r)}(m')\Bigr\}\biggr]\biggr\} \\ - \left(\frac{2\beta}{\lambda} + \frac{\xi}{\gamma'}\right)\log\left(\frac{\eta}{2}\right).$$

Taking $\xi = \frac{\beta\lambda}{2N}$, we get with $\mathbb{P}$ probability at least $1-\eta$

$$\frac{\beta\lambda}{4N}\Bigl(\pi^i_{\exp[-\beta\frac{N}{\lambda}\tanh(\frac{\lambda}{N})R]}\Bigr)^{\otimes 2}\Bigl[\Phi_{\frac{\gamma'}{N}}(M')\Bigr] \\ \leq \log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\Bigl\{\frac{\beta\lambda}{N}\pi^i_{\exp(-\beta r)}(m')\Bigr\}\Bigr]\Bigr\} \\ -\Bigl(\frac{2\beta}{\lambda}+\frac{\beta\lambda}{2N\gamma'}\Bigr)\log\Bigl(\frac{\eta}{2}\Bigr).$$

Putting

$$\lambda = \frac{N^2}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]$$

$$\text{and } \Upsilon(\gamma) \overset{\text{def}}{=} \frac{\gamma\tanh\bigl\{\frac{N}{\gamma}\log\bigl[\cosh(\frac{\gamma}{N})\bigr]\bigr\}}{N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]} \underset{\gamma\to 0}{\sim} 1,$$

this can be rewritten as

$$\frac{\beta N}{4\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\Bigl(\pi^i_{\exp(-\beta\Upsilon(\gamma)R)}\Bigr)^{\otimes 2}\Bigl[\Phi_{\frac{\gamma'}{N}}(M')\Bigr] \\ \leq \log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\Bigl\{\beta\tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\pi^i_{\exp(-\beta r)}(m')\Bigr\}\Bigr]\Bigr\} \\ -\Biggl(\frac{2\beta\gamma}{N^2\log\bigl[\cosh(\frac{\gamma}{N})\bigr]}+\frac{\beta N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]}{2\gamma\gamma'}\Biggr)\log\Bigl(\frac{\eta}{2}\Bigr).$$

It is now tempting to simplify the picture a little bit by setting $\gamma' = \gamma$, leading to

Proposition 2.2.8. *With $\mathbb{P}$ probability at least $1-\eta$, for any $i \in \mathbb{N}$, any $\beta \in \mathbb{R}_+$,*

$$\mathcal{C}\bigl[\pi^i_{\exp(-\beta r)}\bigr] \geq \underline{\mathcal{C}}(i,\beta) \\ \overset{\text{def}}{=} \frac{1}{\zeta-1}\Biggl\{\frac{N}{4}\log\bigl[\cosh(\tfrac{\zeta\beta}{N})\bigr]\Bigl(\pi^i_{\exp(-\beta\Upsilon(\zeta\beta)R)}\Bigr)^{\otimes 2}\Bigl[\Phi_{\frac{\zeta\beta}{N}}(M')\Bigr] \\ +\Biggl(\frac{2\zeta^2\beta^2}{N^2\log\bigl[\cosh(\frac{\zeta\beta}{N})\bigr]}+\frac{N\log\bigl[\cosh(\frac{\zeta\beta}{N})\bigr]}{2\zeta\beta}\Biggr)\log\bigl[2^{-1}\nu(\beta)\mu(i)\eta\bigr] \\ -(\zeta+1)\Bigl\{\log\bigl[\nu(\beta)\mu(i)\bigr]+2^{-1}\log\bigl(3^{-1}\epsilon\bigr)\Bigr\}\Biggr\},$$

where $\mathcal{C}\bigl[\pi^i_{\exp(-\beta r)}\bigr]$ is defined by equation (2.18, page 75).

We are now going to analyse Theorem 2.2.4 (page 72). For this, we will also need an upper bound for $S_\lambda(\rho,\rho')$, defined by equation (2.13, page 70), using M' and empirical complexities, because of the special relations between empirical complexities induced by the selection algorithm. To this purpose, a useful alternative to Proposition 2.2.6 (page 76) is to write, with $\mathbb{P}$ probability at least $1-\eta$,

$$\gamma\rho\otimes\rho'(m') \leq \gamma\rho\otimes\rho'\bigl[\Phi_{-\frac{\gamma}{N}}(M')\bigr] \\ +\mathcal{K}\bigl[\rho,\pi^i_{\exp(-\lambda R)}\bigr]+\mathcal{K}\bigl[\rho',\pi^j_{\exp(-\lambda' R)}\bigr]-\log(\eta),$$

and thus at least with $\mathbb{P}$ probability $1-3\eta$,

$$\gamma\rho\otimes\rho'(m') \leq \gamma\rho\otimes\rho'\bigl[\Phi_{-\frac{\gamma}{N}}(M')\bigr] + (1-\zeta^{-1})^{-1}\Bigl\{\mathcal{K}\bigl[\rho, \pi^i_{\exp(-\beta r)}\bigr] + \log\Bigl\{\pi^i_{\exp(-\beta r)}\Bigl[\exp\bigl\{\tfrac{N}{\zeta}\log\bigl[\cosh(\tfrac{\zeta\beta}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} - \zeta^{-1}\log(\eta)\Bigr\} + (1-\zeta^{-1})^{-1}\Bigl\{\mathcal{K}\bigl[\rho, \pi^j_{\exp(-\beta' r)}\bigr] + \log\Bigl\{\pi^j_{\exp(-\beta' r)}\Bigl[\exp\bigl\{\tfrac{N}{\zeta}\log\bigl[\cosh(\tfrac{\zeta\beta'}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\} - \zeta^{-1}\log(\eta)\Bigr\} - \log(\eta).$$

When $\rho = \pi^i_{\exp(-\beta r)}$ and $\rho' = \pi^j_{\exp(-\beta' r)}$, we get with $\mathbb{P}$ probability at least $1-\eta$, for any β, β', $\gamma \in \mathbb{R}_+$, any i, $j \in \mathbb{N}$,

$$\gamma\rho\otimes\rho'(m') \leq \gamma\rho\otimes\rho'\bigl[\Phi_{-\frac{\gamma}{N}}\bigl[(M')\bigr] + \mathcal{C}(\rho) + \mathcal{C}(\rho') - \frac{\zeta+1}{\zeta-1}\Bigl[\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr].$$

Proposition 2.2.9. *With $\mathbb{P}$ probability at least $1-\eta$, for any $\rho = \pi^i_{\exp(-\beta r)}$, any $\rho' = \pi^j_{\exp(-\beta' r)} \in \mathcal{P}$,*

$$S_\lambda(\rho,\rho') \leq \frac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho\otimes\rho'\bigl[\Phi_{-\frac{\gamma}{N}}(M')\bigr] + \frac{1+\frac{N}{\gamma}\log\bigl[\cosh(\frac{\lambda}{N})\bigr]}{\lambda}\bigl[\mathcal{C}(\rho)+\mathcal{C}(\rho')\bigr] - \frac{(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] + \tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}.$$

In order to analyse Theorem 2.2.4 (page 72), we need to index $\mathcal{P} = \bigl\{\rho_1,\dots,\rho_M\bigr\}$ in order of increasing empirical complexity $\mathcal{C}(\rho)$. To deal in a convenient way with this indexation, we will write $\overline{\mathcal{C}}(i,\beta)$ as $\overline{\mathcal{C}}\bigl[\pi^i_{\exp(-\beta r)}\bigr]$, $\underline{\mathcal{C}}(i,\beta)$ as $\underline{\mathcal{C}}\bigl[\pi^i_{\exp(-\beta r)}\bigr]$, and $\overline{S}\bigl[(i,\beta),(j,\beta')\bigr]$ as $\overline{S}\bigl[\pi^i_{\exp(-\beta r)}, \pi^j_{\exp(-\beta' r)}\bigr]$.

With $\mathbb{P}$ probability at least $1-\epsilon$, when $\widehat{t} \leq j < \widehat{k}$, as we already saw,

$$\rho_{\widehat{k}}(R) \leq \rho_i(R) \leq \rho_j(R) + \inf_{\lambda\in\mathbb{R}_+}\Xi_{\frac{\lambda}{N}}\bigl[2S_\lambda(\rho_j,\rho_i)\bigr],$$

where $i = t(j) < \widehat{t}$. Therefore, with $\mathbb{P}$ probability at least $1-\epsilon-\eta$,

$$\rho_i(R) \leq \rho_j(R) + \inf_{\lambda\in\mathbb{R}_+}\Xi_{\frac{\lambda}{N}}\Biggl\{2\frac{N}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho_j\otimes\rho_i\bigl[\Phi_{-\frac{\gamma}{N}}(M')\bigr] + 4\frac{1+\frac{N}{\gamma}\log\bigl[\cosh(\frac{\lambda}{N})\bigr]}{\lambda}\mathcal{C}(\rho_j) - \frac{(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] + \tfrac{N}{\gamma}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}\Biggr\}.$$

We can now remark that

$$\Xi_a(p+q) \leq \Xi_a(p) + q\Xi_a'(p)q \leq \Xi_a(p) + \Xi_a'(0)q = \Xi_a(p) + \frac{a}{\tanh(a)}q$$

and that

$$\Phi_{-a}(p+q) \leq \Phi_{-a}(p) + \Phi_{-a}'(0)q = \Phi_{-a}(p) + \frac{\exp(a)-1}{a}q.$$

Moreover, assuming as usual without substantial loss of generality that there exists $\widetilde{\theta} \in \arg\min_{\Theta} R$, we can split $M'(\theta,\theta') \leq M'(\theta,\widetilde{\theta}) + M'(\widetilde{\theta},\theta')$. Let us then consider the *expected margin function* defined by

$$\varphi(y) = \sup_{\theta\in\Theta} M'(\theta,\widetilde{\theta}) - yR'(\theta,\widetilde{\theta}), \quad y \in \mathbb{R}_+,$$

and let us write for any $y \in \mathbb{R}_+$,

$$\begin{aligned}\rho_j \otimes \rho_i\bigl[\Phi_{-\frac{\lambda}{N}}(M')\bigr] &\leq \rho_j \otimes \rho_i\bigl\{\Phi_{-\frac{\gamma}{N}}\bigl[M'(.,\widetilde{\theta}) + yR'(.,\widetilde{\theta}) + \varphi(y)\bigr]\bigr\} \\ &\leq \rho_j\bigl\{\Phi_{-\frac{\lambda}{N}}\bigl[M'(.,\widetilde{\theta}) + \varphi(y)\bigr]\bigr\} + \frac{Ny\bigl[\exp(\frac{\gamma}{N})-1\bigr]}{\gamma}\bigl[\rho_i(R) - R(\widetilde{\theta})\bigr]\end{aligned}$$

and

$$\begin{aligned}&\left(1 - \frac{2yN\bigl[\exp(\frac{\gamma}{N})-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right)\bigl[\rho_i(R) - R(\widetilde{\theta})\bigr] \\ &\quad\leq \bigl[\rho_j(R) - R(\widetilde{\theta})\bigr] + \Xi_{\frac{\lambda}{N}}\Biggl\{\frac{2N}{\lambda}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\rho_j\bigl\{\Phi_{-\frac{\gamma}{N}}\bigl[M'(.,\widetilde{\theta}) + \varphi(y)\bigr]\bigr\} \\ &\qquad + 4\frac{1 + \frac{N}{\gamma}\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\lambda}\mathcal{C}(\rho_j) \\ &\qquad - \frac{2(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] + \tfrac{N}{\gamma}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}\Biggr\}.\end{aligned}$$

With $\mathbb{P}$ probability at least $1-\epsilon-\eta$, for any $\lambda, \gamma, x, y \in \mathbb{R}_+$, any $j \in \bigl\{\widehat{t}, \dots, \widehat{k}-1\bigr\}$,

$$\begin{aligned}&\rho_{\widehat{k}}(R) - R(\widetilde{\theta}) \leq \rho_i(R) - R(\widetilde{\theta}) \\ &\quad\leq \left(1 - \frac{2yN\bigl[\exp(\frac{\gamma}{N})-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right)^{-1}\Biggl\{ \\ &\qquad\left(1 + \frac{2xN\bigl[\exp(\frac{\gamma}{N})-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right)\bigl[\rho_j(R) - R(\widetilde{\theta})\bigr] \\ &+ \Xi_{\frac{\lambda}{N}}\Biggl\{\frac{2N}{\lambda}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\Phi_{-\frac{\gamma}{N}}\bigl[\varphi(x) + \varphi(y)\bigr] \\ &\qquad + 4\frac{1 + \frac{N}{\gamma}\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\lambda}\overline{\mathcal{C}}(\rho_j) \\ &\qquad - \frac{2(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] + \tfrac{N}{\gamma}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}\Biggr\}\Biggr\}.\end{aligned}$$

Now we have to get an upper bound for $\rho_j(R)$. We can write $\rho_j = \pi^{\ell}_{\exp(-\beta' r)}$, as we assumed that all the posterior distributions in $\mathcal{P}$ are of this special form. Moreover, we already know from Theorem 2.1.8 (page 58) that with $\mathbb{P}$ probability at least $1-\eta$,

$$\bigl[N\sinh\bigl(\tfrac{\beta'}{N}\bigr) - \beta'\zeta^{-1}\bigr]\bigl[\pi^{\ell}_{\exp(-\beta' r)}(R) - \pi^{\ell}_{\exp(-\beta'\zeta^{-1}R)}(R)\bigr] \\ \leq C^{\ell}(\beta'\zeta^{-1},\beta') - \log\bigl[\nu(\beta')\mu(\ell)\eta\bigr].$$

This proves that with $\mathbb{P}$ probability at least $1-\epsilon-2\eta$,

$$\begin{aligned} \rho_{\widehat{k}}(R) &\leq R(\widetilde{\theta}) \\ &+ \left(1 - \frac{2yN\bigl[\exp\bigl(\frac{\gamma}{N}\bigr)-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right)^{-1}\Biggl\{ \\ &\left(1 + \frac{2xN\bigl[\exp\bigl(\frac{\gamma}{N}\bigr)-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right) \\ &\qquad\times\left(\pi^{\ell}_{\exp(-\zeta^{-1}\beta' R)}(R) - R(\widetilde{\theta}) + \frac{C^{\ell}(\zeta^{-1}\beta',\beta') - \log\bigl[\nu(\beta')\mu(\ell)\eta\bigr]}{N\sinh(\frac{\beta'}{N}) - \zeta^{-1}\beta'}\right) \\ &\qquad\qquad + \Xi_{\frac{\lambda}{N}}\biggl\{\frac{2N}{\lambda}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\Phi_{-\frac{\gamma}{N}}\bigl[\varphi(x)+\varphi(y)\bigr] \\ &\qquad\qquad\qquad + 4\frac{1+\frac{N}{\gamma}\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\lambda}\overline{\mathcal{C}}(\ell,\beta') \\ &\qquad - \frac{2(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] + \tfrac{N}{\gamma}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}\biggr\}\Biggr\}. \end{aligned}$$

The case when $j \in \bigl\{\widehat{k}+1,\dots,M\bigr\}\setminus(\arg\max t)$ is dealt with exactly in the same way, with $i = t(j)$ replaced directly with $\widehat{k}$ itself, leading to the same inequality.

The case when $j \in (\arg\max t)$ is dealt with bounding first $\rho_{\widehat{k}}(R) - R(\widetilde{\theta})$ in terms of $\rho_{\widehat{t}}(R) - R(\widetilde{\theta})$, and this latter in terms of $\rho_j(R) - R(\widetilde{\theta})$. Let us put

$$\begin{aligned} A(\lambda,\gamma) &= \left(1 - \frac{2xN\bigl[\exp\bigl(\frac{\gamma}{N}\bigr)-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}\right), \\ B(\lambda,\gamma) &= 1 + \frac{2yN\bigl[\exp\bigl(\frac{\gamma}{N}\bigr)-1\bigr]\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\gamma\tanh\bigl(\frac{\lambda}{N}\bigr)}, \\ D(\lambda,\gamma,\rho_j) &= \Xi_{\frac{\lambda}{N}}\biggl\{\frac{2N}{\lambda}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\Phi_{-\frac{\gamma}{N}}\bigl[\varphi(x)+\varphi(y)\bigr] \\ &\qquad + 4\frac{1+\frac{N}{\gamma}\log\bigl[\cosh\bigl(\frac{\lambda}{N}\bigr)\bigr]}{\lambda}\overline{\mathcal{C}}(\rho_j) \\ &\quad - \frac{2(\zeta+1)}{(\zeta-1)\lambda}\Bigl\{\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr] \\ &\qquad + \frac{N}{\gamma}\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr]\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr\}\biggr\}, \end{aligned} \tag{2.20}$$

where $\overline{\mathcal{C}}(\rho_j) = \overline{\mathcal{C}}(\ell,\beta')$ is defined, when $\rho_j = \pi^{\ell}_{\exp(-\beta' r)}$, by equation (2.19, page

77). We obtain, still with $\mathbb{P}$ probability $1-\epsilon-2\eta$,

$$\rho_{\widehat{k}}(R) - R(\widetilde{\theta}) \leq \frac{B(\lambda,\gamma)}{A(\lambda,\gamma)}\bigl[\rho_{\widehat{t}}(R) - R(\widetilde{\theta})\bigr] + \frac{D(\lambda,\gamma,\rho_j)}{A(\lambda,\gamma)},$$
$$\rho_{\widehat{t}}(R) - R(\widetilde{\theta}) \leq \frac{B(\lambda,\gamma)}{A(\lambda,\gamma)}\bigl[\rho_j(R) - R(\widetilde{\theta})\bigr] + \frac{D(\lambda,\gamma,\rho_j)}{A(\lambda,\gamma)}.$$

The use of the factor $D(\lambda,\gamma,\rho_j)$ in the first of these two inequalities, instead of $D(\lambda,\gamma,\rho_{\widehat{t}})$, is justified by the fact that $\mathcal{C}(\rho_{\widehat{t}}) \leq \mathcal{C}(\rho_j)$. Combining the two we get

$$\rho_{\widehat{k}}(R) \leq R(\widetilde{\theta}) + \frac{B(\lambda,\gamma)^2}{A(\lambda,\gamma)^2}\bigl[\rho_j(R) - R(\widetilde{\theta})\bigr] + \left[\frac{B(\lambda,\gamma)}{A(\lambda,\gamma)} + 1\right]\frac{D(\lambda,\gamma,\rho_j)}{A(\lambda,\gamma)}.$$

Since it is the worst bound of all cases, it holds for any value of j, proving

Theorem 2.2.10. *With $\mathbb{P}$ probability at least $1-\epsilon-2\eta$,*

$$\rho_{\widehat{k}}(R) \leq R(\widetilde{\theta}) + \inf_{i,\beta,\lambda,\gamma,x,y}\Biggl\{ \frac{B(\lambda,\gamma)^2}{A(\lambda,\gamma)^2}\Bigl[\pi^i_{\exp(-\beta r)}(R) - R(\widetilde{\theta})\Bigr] + \left[\frac{B(\lambda,\gamma)}{A(\lambda,\gamma)} + 1\right]\frac{D(\lambda,\gamma,\pi^i_{\exp(-\beta r)})}{A(\lambda,\gamma)}\Biggr\}$$
$$\leq R(\widetilde{\theta}) + \inf_{i,\beta,\lambda,\gamma,x,y}\Biggl\{ \frac{B(\lambda,\gamma)^2}{A(\lambda,\gamma)^2}\left(\pi^i_{\exp(-\zeta^{-1}\beta R)}(R) - R(\widetilde{\theta}) + \frac{C^i(\zeta^{-1}\beta,\beta) - \log\bigl[\nu(\beta)\mu(i)\eta\bigr]}{N\sinh\bigl(\frac{\beta}{N}\bigr) - \zeta^{-1}\beta}\right) + \left[\frac{B(\lambda,\gamma)}{A(\lambda,\gamma)} + 1\right]\frac{D(\lambda,\gamma,\pi^i_{\exp(-\beta r)})}{A(\lambda,\gamma)}\Biggr\},$$

where the notation $A(\lambda,\gamma)$, $B(\lambda,\gamma)$ and $D(\lambda,\gamma,\rho)$ is defined by equation (2.20 page 83) and where the notation $C^i(\beta,\gamma)$ is defined in Proposition 2.2.5 (page 75).

The bound is a little involved, but as we will prove next, it gives the same rate as Theorem 2.1.15 (page 65) and its corollaries, when we work with a single model (meaning that the support of μ is reduced to one point) and the goal is to choose adaptively the temperature of the Gibbs posterior, except for the appearance of the union bound factor $-\log\bigl[\nu(\beta)\bigr]$ which can be made of order $\log\bigl[\log(N)\bigr]$ without spoiling the order of magnitude of the bound.

We will encompass the case when one must choose between possibly several parametric models. Let us assume that each π^i is supported by some measurable parameter subset Θ_i (meaning that $\pi^i(\Theta_i) = 1$), let us also assume that the behaviour of π^i is parametric in the sense that there exists a dimension $d_i \in \mathbb{R}_+$ such that

$$\sup_{\beta\in\mathbb{R}_+} \beta\bigl[\pi^i_{\exp(-\beta R)}(R) - \inf_{\Theta_i} R\bigr] \leq d_i. \tag{2.21}$$

Then

$$C^i(\lambda,\gamma) \leq \log\Biggl\{\pi^i_{\exp(-\lambda R)}\biggl[\exp\Bigl\{2N\sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 M'(.,\widetilde{\theta})\Bigr\}\biggr]\Biggr\}$$

$$
\begin{aligned}
&+ 2N \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \pi^i_{\exp(-\lambda R)} \bigl[M'(.,\widetilde{\theta})\bigr] \\
&\leq \log\Biggl\{ \pi^i_{\exp(-\lambda R)} \biggl[\exp 2xN \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \bigl[R - R(\widetilde{\theta})\bigr] \biggr\} \biggr] \Biggr\} \\
&\quad + 2xN \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \pi^i_{\exp(-\lambda R)} \bigl[R - R(\widetilde{\theta})\bigr] \\
&\quad + 4N \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \varphi(x) \\
&\leq 2xN \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \pi^i_{\exp\{-[\lambda - 2xN \sinh(\frac{\gamma}{2N})^2] R\}} \bigl[R - R(\widetilde{\theta})\bigr] \\
&\quad + 2xN \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \pi^i_{\exp(-\lambda R)} \bigl[R - R(\widetilde{\theta})\bigr] \\
&\quad + 4N \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \varphi(x).
\end{aligned}
$$

Thus

$$
C^i(\lambda,\gamma) \leq 4N \sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 \Biggl(x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] + \varphi(x) + \frac{xd_i}{2\lambda} + \frac{xd_i}{2\lambda - 4xN\sinh\bigl(\frac{\gamma}{2N}\bigr)^2} \Biggr).
$$

In the same way,

$$
\begin{aligned}
\overline{\mathcal{C}}(i,\beta) \leq \tfrac{8N}{\zeta - 1} \sinh\bigl(\tfrac{\zeta\beta}{2N}\bigr)^2 \Biggl[& x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] + \varphi(x) \\
& + \frac{\zeta x d_i}{2N\sinh\bigl(\frac{\zeta\beta}{N}\bigr)} \Biggl(1 + \frac{1}{1 - x\zeta\tanh\bigl(\frac{\zeta\beta}{2N}\bigr)} \Biggr) \Biggr] \\
+ 2N\Bigl[\exp\Bigl(\tfrac{\zeta^2\beta^2}{2N^2(\zeta-1)}\Bigr) - 1\Bigr] \Biggl(& \varphi(x) + x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] \\
& + \frac{x\zeta d_i}{N\sinh\bigl(\frac{\zeta\beta}{N}\bigr) - x\zeta N\bigl[\exp\bigl(\frac{\zeta^2\beta^2}{2N^2(\zeta-1)}\bigr) - 1\bigr]} \Biggr) \\
& - \frac{(\zeta+1)}{(\zeta-1)} \Bigl[2\log\bigl[\nu(\beta)\mu(i)\bigr] + \log\bigl(\tfrac{\eta}{2}\bigr) \Bigr].
\end{aligned}
$$

In order to keep the right order of magnitude while simplifying the bound, let us consider

$$
\begin{aligned}
(2.22) \quad C_1 = \max\Biggl\{ & \zeta - 1, \Bigl(\tfrac{2N}{\zeta\beta_{\max}}\Bigr)^2 \sinh\Bigl(\tfrac{\zeta\beta_{\max}}{2N}\Bigr)^2, \\
& \tfrac{2N^2(\zeta-1)}{\zeta^2\beta^2_{\max}} \Bigl[\exp\Bigl(\tfrac{\zeta^2\beta^2_{\max}}{2N^2(\zeta-1)}\Bigr) - 1\Bigr] \Biggr\}.
\end{aligned}
$$

Then, for any $\beta \in (0, \beta_{\max})$,

$$
\begin{aligned}
\overline{\mathcal{C}}(i,\beta) \leq \inf_{y \in \mathbb{R}_+} \frac{3C_1\zeta^2\beta^2}{(\zeta-1)N} \Biggl[& y\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] + \varphi(y) + \frac{yd_i}{\beta\bigl[1 - \frac{yC_1\zeta^2\beta}{2(\zeta-1)N}\bigr]} \Biggr] \\
& - \frac{(\zeta+1)}{(\zeta-1)} \Bigl[2\log\bigl[\nu(\beta)\mu(i)\bigr] + \log\bigl(\tfrac{\eta}{2}\bigr) \Bigr].
\end{aligned}
$$

Thus

$$D\bigl[\lambda,\gamma,\pi^i_{\exp(-\beta r)}\bigr] \leq \frac{\lambda}{N\tanh\bigl(\frac{\lambda}{N}\bigr)}\Biggl\{\frac{\lambda\bigl[\exp\bigl(\frac{\gamma}{N}\bigr)-1\bigr]}{\gamma}\bigl[\varphi(x)+\varphi(y)\bigr] \\ + 4\frac{1+\frac{\lambda^2}{2N\gamma}}{\lambda}\Biggl[\frac{3C_1\zeta^2\beta^2}{(\zeta-1)N}\Biggl(z\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]+\varphi(z)+\frac{zd_i}{\beta\bigl[1-\frac{zC_1\zeta^2\beta}{2(\zeta-1)N}\bigr]}\Biggr) \\ - \frac{(\zeta+1)}{(\zeta-1)}\Bigl[2\log\bigl[\nu(\beta)\mu(i)\bigr]+\log\bigl(\tfrac{\eta}{2}\bigr)\Bigr]\Biggr] \\ - \frac{2(\zeta+1)}{(\zeta-1)\lambda}\Bigl[\log\bigl[3^{-1}\nu(\lambda)\epsilon\bigr]+\frac{\lambda^2}{2N\gamma}\log\bigl[3^{-1}\nu(\gamma)\eta\bigr]\Bigr]\Biggr\}$$

If we are not seeking tight constants, we can take for the sake of simplicity $\lambda=\gamma=\beta$, $x=y$ and $\zeta=2$.

Let us put

$$C_2 = \max\Biggl\{C_1, \frac{N\bigl[\exp\bigl(\frac{\beta_{\max}}{N}\bigr)-1\bigr]}{\beta_{\max}}, \\ \frac{2N\log\bigl[\cosh\bigl(\frac{\beta_{\max}}{N}\bigr)\bigr]}{\beta_{\max}\tanh\bigl(\frac{\beta_{\max}}{N}\bigr)}, \frac{\beta_{\max}}{N\tanh\bigl(\frac{\beta_{\max}}{N}\bigr)}\Biggr\}, \tag{2.23}$$

so that

$$A(\beta,\beta)^{-1} \leq \Bigl(1-\frac{C_2x\beta}{N}\Bigr)^{-1},$$
$$B(\beta,\beta) \leq 1+\frac{C_2x\beta}{N},$$

$$D\bigl[\beta,\beta,\pi^i_{\exp(-\beta r)}\bigr] \leq C_2^2\frac{2\beta}{N}\varphi(x) \\ + \Bigl(4+\tfrac{2\beta}{N}\Bigr)\frac{C_2}{\beta}\Biggl[\frac{12C_1\beta^2}{N}\Biggl(z\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]+\varphi(z)+\frac{zd_i}{\beta\bigl[1-\frac{2zC_1\beta}{N}\bigr]}\Biggr) \\ - 6\log\bigl[\nu(\beta)\mu(i)\bigr]-3\log\bigl(\tfrac{\eta}{2}\bigr)\Biggr] \\ - \frac{6C_2}{\beta}\Bigl[\log\bigl[3^{-1}\nu(\beta)\epsilon\bigr]+\frac{\beta}{2N}\log\bigl[3^{-1}\nu(\beta)\eta\bigr]\Bigr]$$

and

$$C^i(\zeta^{-1}\beta,\beta) \leq \frac{C_1\beta^2}{N}\Biggl(x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]+\varphi(x)+\frac{2xd_i}{\beta\bigl[1-\frac{x\beta}{N}\bigr]}\Biggr).$$

This leads to

$$\rho_{\widehat{k}}(R) \leq R(\widetilde{\theta}) + \inf_{i,\beta}\Biggl(\frac{1+\frac{C_2x\beta}{N}}{1-\frac{C_2x\beta}{N}}\Biggr)^2\Biggl\{\frac{2d_i}{\beta}+\inf_{\Theta_i} R - R(\widetilde{\theta}) \\ + \frac{2}{\beta}\Biggl[\frac{C_1\beta^2}{N}\Biggl(x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta}\bigr]+\varphi(x)+\frac{2xd_i}{\beta\bigl(1-\frac{x\beta}{N}\bigr)}\Biggr)$$

$$
\begin{aligned}
& - \log\bigl[\nu(\beta)\mu(i)\eta\bigr]\Bigr]\Biggr\} \\
& + \frac{2}{\Bigl(1-\frac{C_2 x\beta}{N}\Bigr)^2}\Biggl\{ C_2^2 \frac{2\beta}{N}\varphi(x) \\
& + \Bigl(4 + \tfrac{2\beta}{N}\Bigr)\frac{C_2}{\beta}\Biggl[\frac{12 C_1 \beta^2}{N}\biggl(x\bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] + \varphi(x) + \frac{x d_i}{\beta\bigl[1-\frac{2xC_1\beta}{N}\bigr]}\biggr) \\
& - 6\log\bigl[\nu(\beta)\mu(i)\bigr] - 3\log\bigl(\tfrac{\eta}{2}\bigr)\Biggr] \\
& - \frac{6C_2}{\beta}\Bigl[\log\bigl[3^{-1}\nu(\beta)\epsilon\bigr] + \frac{\beta}{2N}\log\bigl[3^{-1}\nu(\beta)\eta\bigr]\Bigr]\Biggr\}.
\end{aligned}
$$

We see in this expression that, in order to balance the various factors depending on x it is advisable to choose x such that

$$
\inf_{\Theta_i} R - R(\widetilde{\theta}) = \frac{\varphi(x)}{x},
$$

as long as $x \leq \frac{N}{4C_2\beta}$.

Following Mammen and Tsybakov, let us assume that the usual margin assumption holds: for some real constants $c > 0$ and $\kappa \geq 1$,

$$
R(\theta) - R(\widetilde{\theta}) \geq c\bigl[D(\theta, \widetilde{\theta})\bigr]^{\kappa}.
$$

As $D(\theta, \widetilde{\theta}\,) \geq M'(\theta, \widetilde{\theta}\,)$, this also implies the weaker assumption

$$
R(\theta) - R(\widetilde{\theta}\,) \geq c\bigl[M'(\theta, \widetilde{\theta})\bigr]^{\kappa}, \qquad \theta \in \Theta,
$$

which we will really need and use. Let us take $\beta_{\max} = N$ and

$$
\nu = \frac{1}{\lceil \log_2(N) \rceil} \sum_{k=1}^{\lceil \log_2(N) \rceil} \delta_{2^k}.
$$

Then, as we have already seen, $\varphi(x) \leq (1-\kappa^{-1})\bigl(\kappa c x\bigr)^{-\frac{1}{\kappa-1}}$. Thus $\varphi(x)/x \leq b x^{-\frac{\kappa}{\kappa-1}}$, where $b = (1-\kappa^{-1})\bigl(\kappa c\bigr)^{-\frac{1}{\kappa-1}}$. Let us choose accordingly

$$
x = \min\Biggl\{ x_1 \overset{\text{def}}{=} \biggl(\frac{\inf_{\Theta_i} R - R(\widetilde{\theta})}{b}\biggr)^{-\frac{\kappa-1}{\kappa}}, x_2 \overset{\text{def}}{=} \frac{N}{4C_2\beta}\Biggr\}.
$$

Using the fact that when $r \in (0, \frac{1}{2})$, $\bigl(\frac{1+r}{1-r}\bigr)^2 \leq 1 + 16r \leq 9$, we get with $\mathbb{P}$ probability at least $1-\epsilon$, for any $\beta \in \operatorname{supp} \nu$, in the case when $x = x_1 \leq x_2$,

$$
\begin{aligned}
\rho_{\widehat{k}}(R) \leq \inf_{\Theta_i} R + 538\, C_2^2 \frac{\beta}{N} b^{\frac{\kappa-1}{\kappa}} \bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]^{\frac{1}{\kappa}} \\
+ \frac{C_2}{\beta}\Bigl[138\, d_i + 166 \log\bigl[1 + \log_2(N)\bigr] - 134 \log\bigl[\mu(i)\bigr] - 102 \log(\epsilon) + 724\Bigr],
\end{aligned}
$$

and in the case when $x = x_2 \leq x_1$,

$$
\begin{aligned}
\rho_{\widehat{k}}(R) \leq \inf_{\Theta_i} R + 68 C_1 \bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr] + 269\, C_2^2 \frac{\beta}{N} \varphi(x) \\
+ \frac{C_2}{\beta} \Bigl[138\, d_i + 166 \log\bigl[1 + \log_2(N)\bigr] - 134 \log\bigl[\mu(i)\bigr] - 102 \log(\epsilon) + 724 \Bigr] \\
\leq \inf_{\Theta_i} R + 541 C_2^2 \frac{\beta}{N} \varphi(x) \\
+ \frac{C_2}{\beta} \Bigl[138\, d_i + 166 \log\bigl[1 + \log_2(N)\bigr] - 134 \log\bigl[\mu(i)\bigr] - 102 \log(\epsilon) + 724 \Bigr].
\end{aligned}
$$

Thus with $\mathbb{P}$ probability at least $1 - \epsilon$,

$$
\begin{aligned}
\rho_{\widehat{k}}(R) \leq \inf_{\Theta_i} R + \inf_{\beta \in (1,N)} 1082\, C_2^2 \frac{\beta}{N} \max \Biggl\{ b^{\frac{\kappa-1}{\kappa}} \bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]^{\frac{1}{\kappa}}, \\
b \Bigl(\frac{4 C_2 \beta}{N} \Bigr)^{\frac{1}{\kappa-1}} \Biggr\} \\
+ \frac{C_2}{\beta} \Bigl[138\, d_i + 166 \log\bigl[1 + \log_2(N)\bigr] \\
- 134 \log\bigl[\mu(i)\bigr] - 102 \log(\epsilon) + 724 \Bigr].
\end{aligned}
$$

THEOREM 2.2.11. *With probability at least $1 - \epsilon$, for any $i \in \mathbb{N}$,*

$$
\begin{aligned}
\rho_{\widehat{k}}(R) \leq \inf_{\Theta_i} R \\
+ \max \left\{ 847 C_2^{\frac{3}{2}} \sqrt{\frac{b^{\frac{\kappa-1}{\kappa}} \bigl[\inf_{\Theta_i} R - R(\widetilde{\theta})\bigr]^{\frac{1}{\kappa}} \Bigl\{ d_i + \log\Bigl(\frac{1+\log_2(N)}{\epsilon \mu(i)}\Bigr) + 5 \Bigr\}}{N}}, \right. \\
\left. 2 C_2 \bigl[1082\, b\bigr]^{\frac{\kappa-1}{2\kappa-1}} 4^{\frac{1}{2\kappa-1}} \left\{ \frac{166 C_2 \Bigl[d_i + \log\Bigl(\frac{1+\log_2(N)}{\epsilon \mu(i)}\Bigr) + 5 \Bigr]}{N} \right\}^{\frac{\kappa}{2\kappa-1}} \right\},
\end{aligned}
$$

where C_2, given by equation (2.23 page 86), will in most cases be close to 1*, and in any case less than* 3.2*.*

This result gives a bound of the same form as that given in Theorem 2.1.15 (page 65) in the special case when there is only one model — that is when μ is a Dirac mass, for instance $\mu(1) = 1$, implying that $R(\widetilde{\theta}_1) - R(\widetilde{\theta}) = 0$. Morover the parametric complexity assumption we made for this theorem, given by equation (2.21 page 84), is weaker than the one used in Theorem 2.1.15 and described by equation (2.8, page 62). When there is more than one model, the bound shows that the estimator makes a trade-off between model accuracy, represented by $\inf_{\Theta_i} R - R(\widetilde{\theta})$, and dimension, represented by d_i, and that for optimal parametric sub-models, meaning those for which $\inf_{\Theta_i} R = \inf_{\Theta} R$, the estimator does at least as well as the minimax optimal convergence speed in the best of these.

Another point is that we obtain more explicit constants than in Theorem 2.1.15. It is also clear that a more careful choice of parameters could have brought some improvement in the value of these constants.

These results show that the selection scheme described in this section is a good candidate to perform temperature selection of a Gibbs posterior distribution built within a single parametric model in a rate optimal way, as well as a proposal with proven performance bound for model selection.

2.3. Two step localization

2.3.1. Two step localization of bounds relative to a Gibbs prior.

Let us reconsider the case where we want to choose adaptively among a family of parametric models. Let us thus assume that the parameter set is a disjoint union of measurable sub-models, so that we can write $\Theta = \sqcup_{m \in M} \Theta_m$, where M is some measurable index set. Let us choose some prior probability distribution on the index set $\mu \in \mathcal{M}_+^1(M)$, and some regular conditional prior distribution $\pi : M \rightarrow \mathcal{M}_+^1(\Theta)$, such that $\pi(i, \Theta_i) = 1$, $i \in M$. Let us then study some arbitrary posterior distributions $\nu : \Omega \rightarrow \mathcal{M}_+^1(M)$ and $\rho : \Omega \times M :\rightarrow \mathcal{M}_+^1(\Theta)$, such that $\rho(\omega, i, \Theta_i) = 1$, $\omega \in \Omega$, $i \in M$. We would like to compare $\nu\rho(R)$ with some doubly localized prior distribution $\mu_{\exp[-\frac{\beta}{1+\zeta_2}\pi_{\exp(-\beta R)}(R)]}\bigl[\pi_{\exp(-\beta R)}\bigr](R)$ (where ζ_2 is a positive parameter to be set as needed later on). To ease notation we will define two prior distributions (one being more precisely a conditional distribution) depending on the positive real parameters β and ζ_2, putting

$$\overline{\pi} = \pi_{\exp(-\beta R)} \text{ and } \overline{\mu} = \mu_{\exp[-\frac{\beta}{1+\zeta_2}\overline{\pi}(R)]}. \tag{2.24}$$

Similarly to Theorem 1.4.3 on page 37 we can write for any positive real constants β and γ

$$\mathbb{P}\biggl\{(\overline{\mu}\,\overline{\pi}) \otimes (\overline{\mu}\,\overline{\pi})\biggl[\exp\Bigl[-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})R'\bigr] \\ - \gamma r' - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\Bigr]\biggr]\biggr\} \leq 1,$$

and deduce, using Lemma 1.1.3 on page 4, that

$$\mathbb{P}\biggl\{\exp\biggl[\sup_{\nu \in \mathcal{M}_+^1(M)}\ \sup_{\rho : M \rightarrow \mathcal{M}_+^1(\Theta)} \Bigl\{-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})(\nu\rho - \overline{\mu}\,\overline{\pi})(R)\bigr] \\ - \gamma(\nu\rho - \overline{\mu}\,\overline{\pi})(r) - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr](\nu\rho)\otimes(\overline{\mu}\,\overline{\pi})(m') \\ - \mathcal{K}(\nu, \overline{\mu}) - \nu\bigl[\mathcal{K}(\rho, \overline{\pi})\bigr]\Bigr\}\biggr]\biggr\} \leq 1. \tag{2.25}$$

This will be our starting point in comparing $\nu\rho(R)$ with $\overline{\mu}\,\overline{\pi}(R)$. However, obtaining an empirical bound will require some supplementary efforts. For each index of the model index set M, we can write in the same way

$$\mathbb{P}\biggl\{\overline{\pi} \otimes \overline{\pi}\biggl[\exp\Bigl[-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})R'\bigr] - \gamma r' - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\Bigr]\biggr]\biggr\} \leq 1.$$

Integrating this inequality with respect to $\overline{\mu}$ and using Fubini's lemma for positive functions, we get

$$\mathbb{P}\biggl\{\overline{\mu}(\overline{\pi} \otimes \overline{\pi})\biggl[\exp\Bigl[-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})R'\bigr] - \gamma r' - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\Bigr]\biggr]\biggr\} \leq 1.$$

Note that $\overline{\mu}(\overline{\pi} \otimes \overline{\pi})$ is a probability measure on $M \times \Theta \times \Theta$, whereas $(\overline{\mu}\,\overline{\pi}) \otimes (\overline{\mu}\,\overline{\pi})$ considered previously is a probability measure on $(M \times \Theta) \times (M \times \Theta)$. We get as previously

$$(2.26)\quad \mathbb{P}\biggl\{\exp\biggl[\sup_{\nu\in\mathcal{M}_+^1(M)}\ \sup_{\rho:M\to\mathcal{M}_+^1(\Theta)}\Bigl\{-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})\nu(\rho-\overline{\pi})(R)\bigr] - \gamma\nu(\rho-\overline{\pi})(r) - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\rho\otimes\overline{\pi})(m') - \mathcal{K}(\nu,\overline{\mu}) - \nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr]\Bigr\}\biggr]\biggr\} \leq 1.$$

Let us finally recall that

$$(2.27)\qquad \mathcal{K}(\nu,\overline{\mu}) = \tfrac{\beta}{1+\zeta_2}(\nu-\overline{\mu})\overline{\pi}(R) + \mathcal{K}(\nu,\mu) - \mathcal{K}(\overline{\mu},\mu),$$

$$(2.28)\qquad \mathcal{K}(\rho,\overline{\pi}) = \beta(\rho-\overline{\pi})(R) + \mathcal{K}(\rho,\pi) - \mathcal{K}(\overline{\pi},\pi).$$

From equations (2.25), (2.26) and (2.28) we deduce

PROPOSITION 2.3.1. *For any positive real constants β, γ and ζ_2, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\nu:\Omega\to\mathcal{M}_+^1(M)$ and any conditional posterior distribution $\rho:\Omega\times M\to\mathcal{M}_+^1(\Theta)$,*

$$-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr] - \beta\nu(\rho-\overline{\pi})(R) \leq \gamma(\nu\rho-\overline{\mu}\,\overline{\pi})(r) + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr](\nu\rho)\otimes(\overline{\mu}\,\overline{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\pi)\bigr] - \nu\bigl[\mathcal{K}(\overline{\pi},\pi)\bigr] + \log\bigl(\tfrac{2}{\epsilon}\bigr).$$

and

$$-N\log\bigl[1-\tanh(\tfrac{\gamma}{N})\nu(\rho-\overline{\pi})(R)\bigr] \leq \gamma\nu(\rho-\overline{\pi})(r) + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\rho\otimes\overline{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr] + \log\bigl(\tfrac{2}{\epsilon}\bigr),$$

where the prior distribution $\overline{\mu}\,\overline{\pi}$ is defined by equation (2.24) *on page 89 and depends on β and ζ_2.*

Let us put for short

$$T = \tanh(\tfrac{\gamma}{N}) \text{ and } C = N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr].$$

We will use an entropy compensation strategy for which we need a couple of entropy bounds. We have according to Proposition 2.3.1, with $\mathbb{P}$ probability at least $1-\epsilon$,

$$\nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr] = \beta\nu(\rho-\overline{\pi})(R) + \nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr] \leq \frac{\beta}{NT}\Bigl[\gamma\nu(\rho-\overline{\pi})(r) + C\nu(\rho\otimes\overline{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr] + \log(\tfrac{2}{\epsilon})\Bigr] + \nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr].$$

Similarly

$$\mathcal{K}(\nu,\overline{\mu}) = \frac{\beta}{1+\zeta_2}(\nu-\overline{\mu})\overline{\pi}(R) + \mathcal{K}(\nu,\mu) - \mathcal{K}(\overline{\mu},\mu)$$
$$\leq \frac{\beta}{(1+\zeta_2)NT}\Bigl[\gamma(\nu-\overline{\mu})\overline{\pi}(r) + C(\nu\overline{\pi})\otimes(\overline{\mu}\,\overline{\pi})(m')$$
$$+ \mathcal{K}(\nu,\overline{\mu}) + \log(\tfrac{2}{\epsilon})\Bigr] + \mathcal{K}(\nu,\mu) - \mathcal{K}(\overline{\mu},\mu).$$

Thus, for any positive real constants β, γ and ζ_i, $i = 1,\dots,5$, with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distributions $\nu,\nu_3 : \Omega \to \mathcal{M}_+^1(\Theta)$, any posterior conditional distributions $\rho,\rho_1,\rho_2,\rho_4,\rho_5 : \Omega\times M \to \mathcal{M}_+^1(\Theta)$,

$$-N\log\bigl[1-T(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr] - \beta\nu(\rho-\overline{\pi})(R)$$
$$\leq \gamma(\nu\rho-\overline{\mu}\,\overline{\pi})(r) + C(\nu\rho)\otimes(\overline{\mu}\,\overline{\pi})(m')$$
$$+ \mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr] + \log(\tfrac{2}{\epsilon}),$$

$$\zeta_1\frac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_1,\overline{\pi})\bigr] \leq \zeta_1\gamma\overline{\mu}(\rho_1-\overline{\pi})(r) + \zeta_1 C\overline{\mu}(\rho_1\otimes\overline{\pi})(m')$$
$$+ \zeta_1\overline{\mu}\bigl[\mathcal{K}(\rho_1,\overline{\pi})\bigr] + \zeta_1\log(\tfrac{2}{\epsilon}) + \zeta_1\frac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_1,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr],$$

$$\zeta_2\frac{NT}{\beta}\nu\bigl[\mathcal{K}(\rho_2,\overline{\pi})\bigr] \leq \zeta_2\gamma\nu(\rho_2-\overline{\pi})(r) + \zeta_2 C\nu(\rho_2\otimes\overline{\pi})(m')$$
$$+ \zeta_2\mathcal{K}(\nu,\overline{\mu}) + \zeta_2\nu\bigl[\mathcal{K}(\rho_2,\overline{\pi})\bigr] + \zeta_2\log(\tfrac{2}{\epsilon})$$
$$+ \zeta_2\frac{NT}{\beta}\nu\bigl[\mathcal{K}(\rho_2,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr],$$

$$\zeta_3(1+\zeta_2)\frac{NT}{\beta}\mathcal{K}(\nu_3,\overline{\mu}) \leq \zeta_3\gamma(\nu_3-\overline{\mu})\overline{\pi}(r)$$
$$+ \zeta_3 C\bigl[(\nu_3\overline{\pi})\otimes(\nu_3\rho_1) + (\nu_3\rho_1)\otimes(\overline{\mu}\,\overline{\pi})\bigr](m') + \zeta_3\mathcal{K}(\nu_3,\overline{\mu}) + \zeta_3\log(\tfrac{2}{\epsilon})$$
$$+ \zeta_3(1+\zeta_2)\frac{NT}{\beta}\bigl[\mathcal{K}(\nu_3,\mu)-\mathcal{K}(\overline{\mu},\mu)\bigr],$$

$$\zeta_4\frac{NT}{\beta}\nu_3\bigl[\mathcal{K}(\rho_4,\overline{\pi})\bigr] \leq \zeta_4\gamma\nu_3(\rho_4-\overline{\pi})(r)$$
$$+ \zeta_4 C\nu_3(\rho_4\otimes\overline{\pi})(m') + \zeta_4\mathcal{K}(\nu_3,\overline{\mu}) + \zeta_4\nu_3\bigl[\mathcal{K}(\rho_4,\overline{\pi})\bigr] + \zeta_4\log(\tfrac{2}{\epsilon})$$
$$+ \zeta_4\frac{NT}{\beta}\nu_3\bigl[\mathcal{K}(\rho_4,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr],$$

$$\zeta_5\frac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_5,\overline{\pi})\bigr] \leq \zeta_5\gamma\overline{\mu}(\rho_5-\overline{\pi})(r) + \zeta_5 C\overline{\mu}(\rho_5\otimes\overline{\pi})(m')$$
$$+ \zeta_5\overline{\mu}\bigl[\mathcal{K}(\rho_5,\overline{\pi})\bigr] + \zeta_5\log(\tfrac{2}{\epsilon}) + \zeta_5\frac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_5,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr].$$

Adding these six inequalities and assuming that

$$\zeta_4 \leq \zeta_3\bigl[(1+\zeta_2)\tfrac{NT}{\beta}-1\bigr], \tag{2.29}$$

we find

$$-N\log\bigl[1-T(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr] - \beta(\nu\rho-\overline{\mu}\,\overline{\pi})(R)$$
$$\leq -N\log\bigl[1-T(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr] - \beta(\nu\rho-\overline{\mu}\,\overline{\pi})(R)$$

$$
\begin{aligned}
&+ \zeta_1\bigl(\tfrac{NT}{\beta} - 1\bigr)\overline{\mu}\bigl[\mathcal{K}(\rho_1,\overline{\pi})\bigr] + \zeta_2\bigl(\tfrac{NT}{\beta} - 1\bigr)\nu\bigl[\mathcal{K}(\rho_2,\overline{\pi})\bigr] \\
&\quad + \bigl[\zeta_3(1+\zeta_2)\tfrac{NT}{\beta} - \zeta_3 - \zeta_4\bigr]\mathcal{K}(\nu_3,\overline{\mu}) \\
&\qquad + \zeta_4\bigl(\tfrac{NT}{\beta} - 1\bigr)\nu_3\bigl[\mathcal{K}(\rho_4,\overline{\pi})\bigr] + \zeta_5\bigl(\tfrac{NT}{\beta} - 1\bigr)\overline{\mu}\bigl[\mathcal{K}(\rho_5,\overline{\pi})\bigr] \\
&\leq \gamma(\nu\rho - \overline{\mu}\,\overline{\pi})(r) + \zeta_1\gamma\overline{\mu}(\rho_1 - \overline{\pi})(r) + \zeta_2\gamma\nu(\rho_2 - \overline{\pi})(r) \\
&\quad + \zeta_3\gamma(\nu_3 - \overline{\mu})\overline{\pi}(r) + \zeta_4\gamma\nu_3(\rho_4 - \overline{\pi})(r) + \zeta_5\gamma\overline{\mu}(\rho_5 - \overline{\pi})(r) \\
&\qquad + C\bigl[(\nu\rho)\otimes(\overline{\mu}\,\overline{\pi}) + \zeta_1\overline{\mu}(\rho_1\otimes\overline{\pi}) + \zeta_2\nu(\rho_2\otimes\overline{\pi}) \\
&+ \zeta_3(\nu_3\overline{\pi})\otimes(\nu_3\rho_1) + \zeta_3(\nu_3\rho_1)\otimes(\overline{\mu}\,\overline{\pi}) \\
&\qquad\qquad + \zeta_4\nu_3(\rho_4\otimes\overline{\pi}) + \zeta_5\overline{\mu}(\rho_5\otimes\overline{\pi})\bigr](m') \\
&+ (1+\zeta_2)\bigl[\mathcal{K}(\nu,\mu) - \mathcal{K}(\overline{\mu},\mu)\bigr] + \nu\bigl[\mathcal{K}(\rho,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr] \\
&\quad + \zeta_1\tfrac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_1,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr] + \zeta_2\tfrac{NT}{\beta}\nu\bigl[\mathcal{K}(\rho_2,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr] \\
&+ \zeta_3(1+\zeta_2)\tfrac{NT}{\beta}\bigl[\mathcal{K}(\nu_3,\mu) - \mathcal{K}(\overline{\mu},\mu)\bigr] + \zeta_4\tfrac{NT}{\beta}\nu_3\bigl[\mathcal{K}(\rho_4,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr] \\
&\quad + \zeta_5\tfrac{NT}{\beta}\overline{\mu}\bigl[\mathcal{K}(\rho_5,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr] + (1+\zeta_1+\zeta_2+\zeta_3+\zeta_4+\zeta_5)\log\bigl(\tfrac{2}{\epsilon}\bigr),
\end{aligned}
$$

where we have also used the fact (concerning the 11th line of the preceding inequalities) that

$$
\begin{aligned}
&-\beta(\nu\rho - \overline{\mu}\,\overline{\pi})(R) + \mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr] \\
&\quad \leq -\beta(\nu\rho - \overline{\mu}\,\overline{\pi})(R) + (1+\zeta_2)\mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\overline{\pi})\bigr] \\
&\qquad = (1+\zeta_2)\bigl[\mathcal{K}(\nu,\mu) - \mathcal{K}(\overline{\mu},\mu)\bigr] + \nu\bigl[\mathcal{K}(\rho,\pi) - \mathcal{K}(\overline{\pi},\pi)\bigr].
\end{aligned}
$$

Let us now apply to $\overline{\pi}$ (we shall later do the same with $\overline{\mu}$) the following inequalities, holding for any random functions of the sample and the parameters $h : \Omega\times\Theta \to \mathbb{R}$ and $g : \Omega\times\Theta\to\mathbb{R}$,

$$
\begin{aligned}
&\overline{\pi}(g-h) - \mathcal{K}(\overline{\pi},\pi) \leq \sup_{\rho:\Omega\times M\to\mathcal{M}_+^1(\Theta)} \rho(g-h) - \mathcal{K}(\rho,\pi) \\
&\quad = \log\bigl\{\pi\bigl[\exp(g-h)\bigr]\bigr\} \\
&\qquad = \log\bigl\{\pi\bigl[\exp(-h)\bigr]\bigr\} + \log\bigl\{\pi_{\exp(-h)}\bigl[\exp(g)\bigr]\bigr\} \\
&\qquad\qquad = -\pi_{\exp(-h)}(h) - \mathcal{K}(\pi_{\exp(-h)},\pi) + \log\bigl\{\pi_{\exp(-h)}\bigl[\exp(g)\bigr]\bigr\}.
\end{aligned}
$$

When h and g are observable, and h is not too far from $\beta r \simeq \beta R$, this gives a way to replace $\overline{\pi}$ with a satisfactory empirical approximation. We will apply this method, choosing ρ_1 and ρ_5 such that $\overline{\mu}\,\overline{\pi}$ is replaced either with $\overline{\mu}\rho_1$, when it comes from the first two inequalities or with $\overline{\mu}\rho_5$ otherwise, choosing ρ_2 such that $\nu\overline{\pi}$ is replaced with $\nu\rho_2$ and ρ_4 such that $\nu_3\overline{\pi}$ is replaced with $\nu_3\rho_4$. We will do so because it leads to a lot of helpful cancellations. For those to happen, we need to choose $\rho_i = \pi_{\exp(-\lambda_i r)}$, $i = 1, 2, 4$, where λ_1, λ_2 and λ_4 are such that

$$(1+\zeta_1)\gamma = \zeta_1\tfrac{NT}{\beta}\lambda_1, \tag{2.30}$$

$$\zeta_2\gamma = \bigl(1+\zeta_2\tfrac{NT}{\beta}\bigr)\lambda_2, \tag{2.31}$$

$$(\zeta_4-\zeta_3)\gamma = \zeta_4\frac{NT}{\beta}\lambda_4, \tag{2.32}$$

$$\zeta_3\gamma = \zeta_5\tfrac{NT}{\beta}\lambda_5, \tag{2.33}$$

and to assume that

$$\zeta_4 > \zeta_3. \tag{2.34}$$

We obtain that with $\mathbb{P}$ probability at least $1-\epsilon$,

$$
\begin{aligned}
&- N\log\bigl[1-T(\mu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr]-\beta(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\\
&\qquad\leq \gamma(\nu\rho-\overline{\mu}\,\rho_1)(r)+\zeta_3\gamma(\nu_3\rho_4-\overline{\mu}\rho_5)(r)\\
&\qquad+\zeta_1\tfrac{NT}{\beta}\overline{\mu}\Biggl\{\log\Biggl[\rho_1\Biggl\{\exp\Biggl[C\tfrac{\beta}{NT\zeta_1}\bigl[\nu\rho+\zeta_1\rho_1\bigr](m')\Biggr]\Biggr\}\Biggr]\Biggr\}\\
&\qquad+\bigl(1+\zeta_2\tfrac{NT}{\beta}\bigr)\nu\Biggl\{\log\Biggl\{\rho_2\Biggl\{\exp\Biggl[\tfrac{C}{1+\zeta_2\frac{NT}{\beta}}\zeta_2\rho_2(m')\Biggr]\Biggr\}\Biggr]\Biggr\}\\
&\qquad+\zeta_4\tfrac{NT}{\beta}\nu_3\Biggl\{\log\Biggl[\rho_4\Biggl\{\exp\Biggl[C\tfrac{\beta}{NT\zeta_4}\bigl[\zeta_3\nu_3\rho_1+\zeta_4\rho_4\bigr](m')\Biggr]\Biggr\}\Biggr]\Biggr\}\\
&\qquad+\zeta_5\tfrac{NT}{\beta}\overline{\mu}\Biggl\{\log\Biggl[\rho_5\Biggl\{\exp\Biggl[C\tfrac{\beta}{NT\zeta_5}\bigl[\zeta_3\nu_3\rho_1+\zeta_5\rho_5\bigr](m')\Biggr]\Biggr\}\Biggr]\Biggr\}\\
&\qquad+(1+\zeta_2)\bigl[\mathcal{K}(\nu,\mu)-\mathcal{K}(\overline{\mu},\mu)\bigr]+\nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\rho_2,\pi)\bigr]\\
&\qquad+\zeta_3(1+\zeta_2)\tfrac{NT}{\beta}\bigl[\mathcal{K}(\nu_3,\mu)-\mathcal{K}(\overline{\mu},\mu)\bigr]\\
&\qquad+\left(1+\sum_{i=1}^{5}\zeta_i\right)\log\bigl(\tfrac{2}{\epsilon}\bigr).
\end{aligned}
$$

In order to obtain more cancellations while replacing $\overline{\mu}$ by some posterior distribution, we will choose the constants such that $\lambda_5=\lambda_4$, which can be done by choosing

$$
\zeta_5=\frac{\zeta_3\zeta_4}{\zeta_4-\zeta_3}. \tag{2.35}
$$

We can now replace $\overline{\mu}$ with $\mu_{\exp-\xi_1\rho_1(r)-\xi_4\rho_4(r)}$, where

$$
\xi_1=\frac{\gamma}{(1+\zeta_2)\bigl(1+\frac{NT}{\beta}\zeta_3\bigr)}, \tag{2.36}
$$

$$
\xi_4=\frac{\gamma\zeta_3}{(1+\zeta_2)\bigl(1+\frac{NT}{\beta}\zeta_3\bigr)}. \tag{2.37}
$$

Choosing moreover $\nu_3=\mu_{\exp-\xi_1\rho_1(r)-\xi_4\rho_4(r)}$, to induce some more cancellations, we get

THEOREM 2.3.2. *Let us use the notation introduced above. For any positive real constants satisfying equations (2.29, page 91), (2.30, page 92), (2.31, page 92), (2.32, page 92), (2.33, page 92), (2.34, page 92), (2.35, page 93), (2.36, page 93), (2.37, page 93), with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\nu:\Omega\to\mathcal{M}_+^1(M)$ and any conditional posterior distribution $\rho:\Omega\times M\to\mathcal{M}_+^1(\Theta)$,*

$$
-N\log\bigl[1-T(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\bigr]-\beta(\nu\rho-\overline{\mu}\,\overline{\pi})(R)\leq B(\nu,\rho,\beta),
$$

$$
\begin{aligned}
&\textit{where } B(\nu,\rho,\beta)\overset{\text{def}}{=}\gamma(\nu\rho-\nu_3\rho_1)(r)\\
&\quad+(1+\zeta_2)\bigl(1+\tfrac{NT}{\beta}\zeta_3\bigr)\\
&\qquad\times\log\Biggl\{\nu_3\Biggl[\rho_1\Biggl\{\exp\Biggl[C\tfrac{\beta}{NT\zeta_1}\bigl[\nu\rho+\zeta_1\rho_1\bigr](m')\Biggr]\Biggr\}^{\frac{\zeta_1 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}}
\end{aligned}
$$

$$\times \rho_4 \Biggl\{ \exp \Bigl[C \tfrac{\beta}{NT\zeta_5} \bigl[\zeta_3 \nu_3 \rho_1 + \zeta_5 \rho_4 \bigr](m') \Bigr] \Biggr\}^{\frac{\zeta_5 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}} \Biggr] \Biggr\}$$

$$+ \bigl(1 + \zeta_2 \tfrac{NT}{\beta}\bigr) \nu \Biggl\{ \log \Biggl\{ \rho_2 \Biggl\{ \exp \Bigl[\tfrac{C}{1+\zeta_2 \frac{NT}{\beta}} \zeta_2 \rho_2(m') \Bigr] \Biggr\} \Biggr] \Biggr\}$$

$$+ \zeta_4 \tfrac{NT}{\beta} \nu_3 \Biggl\{ \log \Biggl[\rho_4 \Biggl\{ \exp \Bigl[C \tfrac{\beta}{NT\zeta_4} \bigl[\zeta_3 \nu_3 \rho_1 + \zeta_4 \rho_4 \bigr](m') \Bigr] \Biggr\} \Biggr] \Biggr\}$$

$$+ (1+\zeta_2) \bigl[\mathcal{K}(\nu, \mu) - \mathcal{K}(\nu_3, \mu) \bigr]$$

$$+ \nu \bigl[\mathcal{K}(\rho, \pi) - \mathcal{K}(\rho_2, \pi) \bigr] + \Biggl(1 + \sum_{i=1}^{5} \zeta_i \Biggr) \log\bigl(\tfrac{2}{\epsilon}\bigr).$$

This theorem can be used to find the largest value $\widehat{\beta}(\nu\rho)$ of β such that $B(\nu, \rho, \beta) \leq 0$, thus providing an estimator for $\beta(\nu\rho)$ defined as $\nu\rho(R) = \overline{\mu}_{\beta(\nu\rho)}\overline{\pi}_{\beta(\nu\rho)}(R)$, where we have mentioned explicitly the dependence of $\overline{\mu}$ and $\overline{\pi}$ in β, the constant ζ_2 staying fixed. The posterior distribution $\nu\rho$ may then be chosen to maximize $\widehat{\beta}(\nu\rho)$ within some manageable subset of posterior distributions $\mathcal{P}$, thus gaining the assurance that $\nu\rho(R) \leq \overline{\mu}_{\widehat{\beta}(\nu\rho)}\overline{\pi}_{\widehat{\beta}(\nu\rho)}(R)$, with the largest parameter $\widehat{\beta}(\nu\rho)$ that this approach can provide. Maximizing $\widehat{\beta}(\nu\rho)$ is supported by the fact that $\lim_{\beta \to +\infty} \overline{\mu}_\beta \overline{\pi}_\beta(R) = \operatorname{ess\,inf}_{\mu\pi} R$. Anyhow, there is no assurance (to our knowledge) that $\beta \mapsto \overline{\mu}_\beta \overline{\pi}_\beta(R)$ will be a decreasing function of β all the way, although this may be expected to be the case in many practical situations.

We can make the bound more explicit in several ways. One point of view is to put forward the optimal values of ρ and ν. We can thus remark that

$$\nu\bigl[\gamma\rho(r) + \mathcal{K}(\rho, \pi) - \mathcal{K}(\rho_2, \pi)\bigr] + (1+\zeta_2)\mathcal{K}(\nu, \mu)$$

$$= \nu \Biggl[\mathcal{K}\bigl[\rho, \pi_{\exp(-\gamma r)}\bigr] + \lambda_2 \rho_2(r) + \int_{\lambda^2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha \Biggr] + (1+\zeta_2)\mathcal{K}(\nu, \mu)$$

$$= \nu\bigl\{\mathcal{K}\bigl[\rho, \pi_{\exp(-\gamma r)}\bigr]\bigr\} + (1+\zeta_2)\mathcal{K}\bigl[\nu, \mu_{\exp\left(-\frac{\lambda_2 \rho_2(r)}{1+\zeta_2} - \frac{1}{1+\zeta_2}\int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha\right)}\bigr]$$

$$- (1+\zeta_2) \log \Biggl\{ \mu \Biggl[\exp \Biggl\{ -\frac{\lambda_2}{1+\zeta_2} \rho_2(r) - \frac{1}{1+\zeta_2} \int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha \Biggr\} \Biggr] \Biggr\}.$$

Thus

$$B(\nu, \rho, \beta) = (1+\zeta_2) \Bigl[\xi_1 \nu_3 \rho_1(r) + \xi_4 \nu_3 \rho_4(r)$$

$$+ \log\bigl\{\mu\bigl[\exp\bigl(-\xi_1 \rho_1(r) - \xi_4 \rho_4(r)\bigr)\bigr]\bigr\} \Bigr]$$

$$- (1+\zeta_2) \log \Biggl\{ \mu \Biggl[\exp \Biggl\{ -\frac{\lambda_2}{1+\zeta_2} \rho_2(r) - \frac{1}{1+\zeta_2} \int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha \Biggr\} \Biggr] \Biggr\}$$

$$- \gamma \nu_3 \rho_1(r) + (1+\zeta_2)\bigl(1 + \tfrac{NT}{\beta}\zeta_3\bigr)$$

$$\times \log \Biggl\{ \nu_3 \Biggl[\rho_1 \Biggl\{ \exp \Bigl[C \tfrac{\beta}{NT\zeta_1} \bigl[\nu\rho + \zeta_1 \rho_1 \bigr](m') \Bigr] \Biggr\}^{\frac{\zeta_1 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}}$$

$$\times \rho_4 \Biggl\{ \exp \Bigl[C \tfrac{\beta}{NT\zeta_5} \bigl[\zeta_3 \nu_3 \rho_1 + \zeta_5 \rho_4 \bigr](m') \Bigr] \Biggr\}^{\frac{\zeta_5 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}} \Biggr] \Biggr\}$$

$$
\begin{aligned}
&\qquad + \bigl(1+\zeta_2 \tfrac{NT}{\beta}\bigr)\nu \biggl\{ \log \biggl\{ \rho_2 \biggl\{ \exp\Bigl[\tfrac{C}{1+\zeta_2\frac{NT}{\beta}} \zeta_2 \rho_2(m') \Bigr] \biggr\} \biggr] \biggr\} \\
&\qquad + \zeta_4 \tfrac{NT}{\beta} \nu_3 \biggl\{ \log \biggl[\rho_4 \biggl\{ \exp\Bigl[C \tfrac{\beta}{NT\zeta_4} \bigl[\zeta_3 \nu_3 \rho_1 + \zeta_4 \rho_4 \bigr](m') \Bigr] \biggr\} \biggr] \biggr\} \\
&+ \nu \bigl\{ \mathcal{K}\bigl[\rho, \pi_{\exp(-\gamma r)}\bigr] \bigr\} \\
&\qquad + (1+\zeta_2) \mathcal{K}\Bigl[\nu, \mu_{\exp\bigl(-\frac{\lambda_2 \rho_2(r)}{1+\zeta_2} - \frac{1}{1+\zeta_2}\int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha\bigr)}\Bigr] \\
&\qquad\qquad + \biggl(1 + \sum_{i=1}^{5} \zeta_i\biggr) \log\bigl(\tfrac{2}{\epsilon}\bigr).
\end{aligned}
$$

This formula is better understood when thinking about the following upper bound for the two first lines in the expression of $B(\nu, \rho, \beta)$:

$$
\begin{aligned}
&(1+\zeta_2)\Bigl[\xi_1 \nu_3 \rho_1(r) + \xi_4 \nu_3 \rho_4(r) + \log\bigl\{\mu\bigl[\exp\bigl(-\xi_1\rho_1(r) - \xi_4 \rho_4(r)\bigr)\bigr]\bigr\}\Bigr] \\
&\quad - (1+\zeta_2)\log\biggl\{\mu\biggl[\exp\biggl\{-\frac{\lambda_2}{1+\zeta_2}\rho_2(r) \\
&\qquad\qquad - \frac{1}{1+\zeta_2}\int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha\biggr\}\biggr]\biggr\} - \gamma \nu_3 \rho_1(r) \\
&\qquad\qquad \le \nu_3 \biggl[\lambda_2 \rho_2(r) + \int_{\lambda_2}^{\gamma} \pi_{\exp(-\alpha r)}(r) d\alpha - \gamma \rho_1(r)\biggr].
\end{aligned}
$$

Another approach to understanding Theorem 2.3.2 is to put forward $\rho_0 = \pi_{\exp(-\lambda_0 r)}$, for some positive real constant $\lambda_0 < \gamma$, noticing that

$$
\nu\bigl[\mathcal{K}(\rho_0, \pi) - \mathcal{K}(\rho_2, \pi)\bigr] = \lambda_0 \nu(\rho_2 - \rho_0)(r) - \nu\bigl[\mathcal{K}(\rho_2, \rho_0)\bigr].
$$

Thus

$$
\begin{aligned}
&B(\nu, \rho_0, \beta) \le \nu_3\bigl[(\gamma - \lambda_0)(\rho_0 - \rho_1)(r) + \lambda_0(\rho_2 - \rho_1)(r)\bigr] \\
&\quad + (1+\zeta_2)\bigl(1 + \tfrac{NT}{\beta}\zeta_3\bigr) \\
&\qquad \times \log\biggl\{\nu_3 \biggl[\rho_1\biggl\{\exp\Bigl[C \tfrac{\beta}{NT\zeta_1}\bigl[\nu\rho_0 + \zeta_1 \rho_1\bigr](m')\Bigr]\biggr\}^{\frac{\zeta_1 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}} \\
&\qquad\qquad \times \rho_4\biggl\{\exp\Bigl[C\tfrac{\beta}{NT\zeta_5}\bigl[\zeta_3 \nu_3 \rho_1 + \zeta_5 \rho_4\bigr](m')\Bigr]\biggr\}^{\frac{\zeta_5 NT}{\beta(1+\zeta_2)(1+\frac{NT}{\beta}\zeta_3)}} \biggr]\biggr\} \\
&\qquad + \bigl(1+\zeta_2 \tfrac{NT}{\beta}\bigr)\nu \biggl\{ \log \biggl\{ \rho_2 \biggl\{ \exp\Bigl[\tfrac{C}{1+\zeta_2\frac{NT}{\beta}} \zeta_2 \rho_2(m') \Bigr] \biggr\} \biggr] \biggr\} \\
&\qquad + \zeta_4 \tfrac{NT}{\beta} \nu_3 \biggl\{ \log \biggl[\rho_4 \biggl\{ \exp\Bigl[C \tfrac{\beta}{NT\zeta_4} \bigl[\zeta_3 \nu_3 \rho_1 + \zeta_4 \rho_4 \bigr](m') \Bigr] \biggr\} \biggr] \biggr\} \\
&\quad + (1+\zeta_2)\mathcal{K}\Bigl[\nu, \mu_{\exp\bigl(-\frac{(\gamma - \lambda_0)\rho_0(r) + \lambda_0 \rho_2(r)}{1+\zeta_2}\bigr)}\Bigr] \\
&\qquad\qquad - \nu\bigl[\mathcal{K}(\rho_2, \rho_0)\bigr] + \biggl(1 + \sum_{i=1}^{5}\zeta_i\biggr)\log\bigl(\tfrac{2}{\epsilon}\bigr).
\end{aligned}
$$

In the case when we want to select a single model $\widehat{m}(\omega)$, and therefore to set $\nu = \delta_{\widehat{m}}$, the previous inequality engages us to take
$$\widehat{m} \in \arg\min_{m \in M} (\gamma - \lambda_0)\rho_0(m, r) + \lambda_0 \rho_2(m, r).$$
In parametric situations where

$$\pi_{\exp(-\lambda r)}(r) \simeq r^\star(m) + \frac{d_e(m)}{\lambda},$$

we get

$$(\gamma - \lambda_0)\rho_0(m, r) - \lambda_0 \rho_2(m, r) \simeq \gamma\bigl[r^\star(m) + d_e(m)\bigl(\tfrac{1}{\lambda_0} + \tfrac{\lambda_0 - \lambda_2}{\gamma \lambda_2}\bigr)\bigr],$$

resulting in a linear penalization of the empirical dimension of the models.

2.3.2. Analysis of two step bounds relative to a Gibbs prior. We will not state a formal result, but will nevertheless give some hints about how to establish one. This is a rather technical section, which can be skipped at a first reading , since it will not be used below. We should start from Theorem 1.4.2 (page 35), which gives a deterministic variance term. From Theorem 1.4.2, after a change of prior distribution, we obtain for any positive constants α_1 and α_2, any prior distributions $\widetilde{\mu}_1$ and $\widetilde{\mu}_2 \in \mathcal{M}_+^1(M)$, for any prior conditional distributions $\widetilde{\pi}_1$ and $\widetilde{\pi}_2 : M \rightarrow \mathcal{M}_+^1(\Theta)$, with $\mathbb{P}$ probability at least $1 - \eta$, for any posterior distributions $\nu_1 \rho_1$ and $\nu_2 \rho_2$,

$$\begin{aligned}\alpha_1(\nu_1\rho_1 - \nu_2\rho_2)(R) \leq \alpha_2(\nu_1\rho_1 - \nu_2\rho_2)(r)& \\ + \mathcal{K}\bigl[(\nu_1\rho_1) \otimes (\nu_2\rho_2), (\widetilde{\mu}_1 \,\widetilde{\pi}_1) \otimes (\widetilde{\mu}_2 \,\widetilde{\pi}_2)\bigr]& \\ + \log\Bigl\{(\widetilde{\mu}_1 \,\widetilde{\pi}_1) \otimes (\widetilde{\mu}_2 \,\widetilde{\pi}_2)\Bigl[\exp\bigl\{-\alpha_2 \Psi_{\frac{\alpha_2}{N}}(R', M') + \alpha_1 R'\bigr\}\Bigr]\Bigr\} - \log(\eta).&\end{aligned}$$

Applying this to $\alpha_1 = 0$, we get that

$$\begin{aligned}(\nu\rho - \nu_3\rho_1)(r) \leq \frac{1}{\alpha_2}\Bigl[&\mathcal{K}\bigl[(\nu\rho) \otimes (\nu_3\rho_1), (\widetilde{\mu}\,\widetilde{\pi}) \otimes (\widetilde{\mu}_3 \,\widetilde{\pi}_1)\bigr] \\ &+ \log\Bigl\{(\widetilde{\mu}\,\widetilde{\nu}) \otimes (\widetilde{\mu}_3 \,\widetilde{\pi}_1)\Bigl[\exp\bigl\{\alpha_2 \Psi_{-\frac{\alpha_2}{N}}(R', M')\bigr\}\Bigr]\Bigr\} - \log(\eta)\Bigr].\end{aligned}$$

In the same way, to bound quantities of the form

$$\begin{aligned}\log\Biggl\{\nu_3\Biggl[\rho_1\Bigl\{\exp\Bigl[C_1(\nu\rho + \zeta_1\rho_1)(m')\Bigr]\Bigr\}^{p_1}& \\ \times \rho_4\Bigl\{\exp\Bigl[C_2\bigl[\zeta_3\nu_3\rho_1 + \zeta_5\rho_4\bigr](m')\Bigr]\Bigr\}^{p_2}\Biggr]\Biggr\}& \\ = \sup_{\nu_5}\Biggl\{p_1 \sup_{\rho_5}\Bigl\{C_1\bigl[(\nu\rho) \otimes (\nu_5\rho_5) + \zeta_1\nu_5(\rho_1 \otimes \rho_5)\bigr](m') - \mathcal{K}(\rho_5, \rho_1)\Bigr\}& \\ + p_2 \sup_{\rho_6}\Bigl\{C_2\bigl[\zeta_3(\nu_3\rho_1) \otimes (\nu_5\rho_6)& \\ + \zeta_5\nu_5(\rho_4 \otimes \rho_6)\bigr](m') - \mathcal{K}(\rho_6, \rho_4)\Bigr\} - \mathcal{K}(\nu_5, \nu_3)\Biggr\},&\end{aligned}$$

where C_1, C_2, p_1 and p_2 are positive constants, and similar terms, we need to use inequalities of the type: for any prior distributions $\widetilde{\mu}_i\,\widetilde{\pi}_i$, $i=1,2$, with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distributions $\nu_i\rho_i$, $i=1,2$,

$$\alpha_3(\nu_1\rho_1)\otimes(\nu_2\rho_2)(m') \leq \log\Bigl\{(\widetilde{\mu}_1\,\widetilde{\pi}_1)\otimes(\widetilde{\mu}_2\,\widetilde{\pi}_2)\exp\Bigl[\alpha_3\Phi_{\frac{-\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}\bigl[(\nu_1\rho_1)\otimes(\nu_2\rho_2),(\widetilde{\mu}_1\,\widetilde{\pi}_1)\otimes(\widetilde{\mu}_2\,\widetilde{\pi}_2)\bigr] - \log(\eta).$$

We need also the variant: with $\mathbb{P}$ probability at least $1-\eta$, for any posterior distribution $\nu_1:\Omega\rightarrow\mathcal{M}_+^1(M)$ and any conditional posterior distributions $\rho_1,\rho_2:\Omega\times M\rightarrow\mathcal{M}_+^1(\Theta)$,

$$\alpha_3\nu_1(\rho_1\otimes\rho_2)(m') \leq \log\Bigl\{\widetilde{\mu}_1\bigl(\widetilde{\pi}_1\otimes\widetilde{\pi}_2\bigr)\exp\Bigl[\alpha_3\Phi_{-\frac{\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}(\nu_1,\widetilde{\mu}_1) + \nu_1\bigl\{\mathcal{K}\bigl[\rho_1\otimes\rho_2,\widetilde{\pi}_1\otimes\widetilde{\pi}_2\bigr]\bigr\} - \log(\eta).$$

We deduce that

$$\log\Biggl\{\nu_3\Biggl[\rho_1\Biggl\{\exp\Bigl[C_1(\nu\rho+\zeta_1\rho_1)(m')\Bigr]\Biggr\}^{p_1} \\ \times\rho_4\Biggl\{\exp\Bigl[C_2\bigl[\zeta_3\nu_3\rho_1+\zeta_5\rho_4\bigr](m')\Bigr]\Biggr\}^{p_2}\Biggr]\Biggr\} \\ \leq \sup_{\nu_5}\Biggl\{p_1\sup_{\rho_5}\Biggl[\frac{C_1}{\alpha_3}\Biggl\{\log\Bigl\{(\widetilde{\mu}\,\widetilde{\pi})\otimes(\widetilde{\mu}_5\,\widetilde{\pi}_5)\exp\Bigl[\alpha_3\Phi_{-\frac{\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}\bigl[(\nu\rho)\otimes(\nu_5\rho_5),(\widetilde{\mu}\,\widetilde{\pi}\otimes(\widetilde{\mu}_5\,\widetilde{\pi}_5)\bigr] + \log\bigl(\tfrac{2}{\eta}\bigr) \\ + \zeta_1\Biggl[\log\Bigl\{\widetilde{\mu}_5\bigl(\widetilde{\pi}_1\otimes\widetilde{\pi}_5\bigr)\exp\Bigl[\alpha_3\Phi_{-\frac{\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}(\nu_5,\widetilde{\mu}_5) + \nu_5\bigl\{\mathcal{K}\bigl[\rho_1\otimes\rho_5,\widetilde{\pi}_1\otimes\widetilde{\pi}_5\bigr]\bigr\} + \log\bigl(\tfrac{2}{\eta}\bigr)\Biggr]\Biggr\} - \mathcal{K}(\rho_5,\rho_1)\Biggr] \\ + p_2\sup_{\rho_6}\Biggl[\frac{C_1}{\alpha_3}\Biggl\{\log\Bigl\{(\widetilde{\mu}_3\,\widetilde{\pi}_1)\otimes(\widetilde{\mu}_5\,\widetilde{\pi}_6)\exp\Bigl[\alpha_3\Phi_{-\frac{\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}\bigl[(\nu_3\rho_1)\otimes(\nu_5\rho_6),(\widetilde{\mu}_3\,\widetilde{\pi}_1\otimes(\widetilde{\mu}_5\,\widetilde{\pi}_6)\bigr] + \log\bigl(\tfrac{2}{\eta}\bigr) \\ + \zeta_1\Biggl[\log\Bigl\{\widetilde{\mu}_5\bigl(\widetilde{\pi}_4\otimes\widetilde{\pi}_6\bigr)\exp\Bigl[\alpha_3\Phi_{-\frac{\alpha_3}{N}}(M')\Bigr]\Bigr\} \\ + \mathcal{K}(\nu_5,\widetilde{\mu}_5) + \nu_5\bigl\{\mathcal{K}\bigl[\rho_4\otimes\rho_6,\widetilde{\pi}_4\otimes\widetilde{\pi}_6\bigr]\bigr\} + \log\bigl(\tfrac{2}{\eta}\bigr)\Biggr]\Biggr\} \\ - \mathcal{K}(\rho_6,\rho_4)\Biggr] - \mathcal{K}(\nu_5,\nu_3)\Biggr\}.$$

We are then left with the need to bound entropy terms like $\mathcal{K}(\nu_3\rho_1,\widetilde{\mu}_3\widetilde{\pi}_1)$, where we have the choice of $\widetilde{\mu}_3$ and $\widetilde{\pi}_1$, to obtain a useful bound. As could be expected, we decompose it into

$$\mathcal{K}(\nu_3\rho_1,\widetilde{\mu}_3\widetilde{\pi}_1) = \mathcal{K}(\nu_3,\widetilde{\mu}_3) + \nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr].$$

Let us look after the second term first, choosing $\widetilde{\pi}_1 = \pi_{\exp(-\beta_1 R)}$:

$$\begin{aligned}
\nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr] &= \nu_3\bigl[\beta_1(\rho_1-\widetilde{\pi}_1)(R)+\mathcal{K}(\rho_1,\pi)-\mathcal{K}(\widetilde{\pi}_1,\pi)\bigr]\\
&\leq \frac{\beta_1}{\alpha_1}\biggl[\alpha_2\nu_3(\rho_1-\widetilde{\pi}_1)(r)+\mathcal{K}(\nu_3,\widetilde{\mu}_3)+\nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr]\\
&\quad+\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\alpha_2\Psi_{\frac{\alpha_2}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr]\\
&\quad+\nu_3\bigl[\mathcal{K}(\rho_1,\pi)-\mathcal{K}(\widetilde{\pi}_1,\pi)\bigr]\\
\leq \frac{\beta_1}{\alpha_1}&\biggl[\mathcal{K}(\nu_3,\widetilde{\mu}_3)+\nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr]\\
&\quad+\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\alpha_2\Psi_{\frac{\alpha_2}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr]\\
&\quad+\nu_3\bigl\{\mathcal{K}\bigl[\rho_1,\pi_{\exp(-\frac{\beta_1\alpha_2}{\alpha_1}r)}\bigr]\bigr\}.
\end{aligned}$$

Thus, when the constraint $\lambda_1=\frac{\beta_1\alpha_2}{\alpha_1}$ is satisfied,

$$\begin{aligned}
\nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr] \leq \Bigl(1-\frac{\beta_1}{\alpha_1}\Bigr)^{-1}\frac{\beta_1}{\alpha_1}\biggl[\mathcal{K}(\nu_3,\widetilde{\mu}_3)\\
+\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\alpha_2\Psi_{\frac{\alpha_2}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr].
\end{aligned}$$

We can further specialize the constants, choosing $\alpha_1=N\sinh(\frac{\alpha_2}{N})$, so that

$$-\alpha_2\Psi_{\frac{\alpha_2}{N}}(R',M')+\alpha_1 R'\leq 2N\sinh\Bigl(\frac{\alpha_2}{2N}\Bigr)^2 M'.$$

We can for instance choose $\alpha_2=\gamma$, $\alpha_1=N\sinh(\frac{\gamma}{N})$ and $\beta_1=\lambda_1\frac{N}{\gamma}\sinh(\frac{\gamma}{N})$, leading to

Proposition 2.3.3. *With the notation of Theorem 2.3.2, the constants being set as explained above, putting* $\widetilde{\pi}_1=\pi_{\exp(-\lambda_1\frac{N}{\gamma}\sinh(\frac{\gamma}{N})R)}$, *with* $\mathbb{P}$ *probability at least* $1-\eta$,

$$\begin{aligned}
\nu_3\bigl[\mathcal{K}(\rho_1,\widetilde{\pi}_1)\bigr] \leq \Bigl(1-\frac{\lambda_1}{\gamma}\Bigr)^{-1}\frac{\lambda_1}{\gamma}\biggl[\mathcal{K}(\nu_3,\widetilde{\mu}_3)\\
+\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{2N\sinh(\tfrac{\gamma}{2N})^2 M'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr].
\end{aligned}$$

More generally

$$\begin{aligned}
\nu_3\bigl[\mathcal{K}(\rho,\widetilde{\pi}_1)\bigr] \leq \Bigl(1-\frac{\lambda_1}{\gamma}\Bigr)^{-1}\frac{\lambda_1}{\gamma}\biggl[\mathcal{K}(\nu_3,\widetilde{\mu}_3)\\
+\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{2N\sinh(\tfrac{\gamma}{2N})^2 M'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr]\\
+\Bigl(1-\frac{\lambda_1}{\gamma}\Bigr)^{-1}\nu_3\bigl[\mathcal{K}(\rho,\rho_1)\bigr].
\end{aligned}$$

In a similar way, let us now choose $\widetilde{\mu}_3=\mu_{\exp[-\alpha_3\overline{\pi}(R)]}$. We can write

$$\mathcal{K}(\nu,\widetilde{\mu}_3)=\alpha_3(\nu-\widetilde{\mu}_3)\overline{\pi}(R)+\mathcal{K}(\nu,\mu)-\mathcal{K}(\widetilde{\mu}_3,\mu)$$

$$
\begin{aligned}
&\leq \frac{\alpha_3}{\alpha_1}\biggl[\alpha_2(\nu-\widetilde{\mu}_3)\overline{\pi}(r)+\mathcal{K}(\nu,\widetilde{\mu}_3) \\
&+\log\Bigl\{(\widetilde{\mu}_3\overline{\pi})\otimes(\widetilde{\mu}_3\overline{\pi})\Bigl[\exp\bigl\{-\alpha_2\Psi_{\frac{\alpha_2}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}-\log(\eta)\biggr] \\
&+\mathcal{K}(\nu,\mu)-\mathcal{K}(\widetilde{\mu}_3,\mu).
\end{aligned}
$$

Let us choose $\alpha_2=\gamma$, $\alpha_1=N\sinh(\frac{\gamma}{N})$, and let us add some other entropy inequalities to get rid of $\overline{\pi}$ in a suitable way, the approach of entropy compensation being the same as that used to obtain the empirical bound of Theorem 2.3.2 (page 93). This results with $\mathbb{P}$ probability at least $1-\eta$ in

$$
\begin{aligned}
\Bigl(1-\frac{\alpha_3}{\alpha_1}\Bigr)\mathcal{K}(\nu,\widetilde{\mu}_3) &\leq \frac{\alpha_3}{\alpha_1}\biggl[\gamma(\nu-\widetilde{\mu}_3)\overline{\pi}(r) \\
&+\log\Bigl\{(\widetilde{\mu}_3\overline{\pi})\otimes(\widetilde{\mu}_3\overline{\pi})\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}+\log(\tfrac{2}{\eta})\biggr] \\
&+\mathcal{K}(\nu,\mu)-\mathcal{K}(\widetilde{\mu}_3,\mu), \\
\zeta_6\Bigl(1-\frac{\beta}{\alpha_1}\Bigr)\widetilde{\mu}_3\bigl[\mathcal{K}(\rho_6,\overline{\pi})\bigr] &\leq \zeta_6\frac{\beta}{\alpha_1}\biggl[\gamma\widetilde{\mu}_3(\rho_6-\overline{\pi})(r) \\
&+\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}+\log(\tfrac{2}{\eta})\biggr] \\
&+\zeta_6\widetilde{\mu}_3\bigl[\mathcal{K}(\rho_6,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr], \\
\zeta_7\Bigl(1-\frac{\beta}{\alpha_1}\Bigr)\widetilde{\mu}_3\bigl[\mathcal{K}(\rho_7,\overline{\pi})\bigr] &\leq \zeta_7\frac{\beta}{\alpha_1}\biggl[\gamma\widetilde{\mu}_3(\rho_7-\overline{\pi})(r) \\
&+\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}+\log(\tfrac{2}{\eta})\biggr] \\
&+\zeta_7\widetilde{\mu}_3\bigl[\mathcal{K}(\rho_7,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr], \\
\zeta_8\Bigl(1-\frac{\beta}{\alpha_1}\Bigr)\nu\bigl[\mathcal{K}(\rho_8,\overline{\pi})\bigr] &\leq \zeta_8\frac{\beta}{\alpha_1}\biggl[\gamma\nu(\rho_8-\overline{\pi})(r)+\mathcal{K}(\nu,\widetilde{\mu}_3) \\
&+\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}+\log(\tfrac{2}{\eta})\biggr] \\
&+\zeta_8\nu\bigl[\mathcal{K}(\rho_8,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr], \\
\zeta_9\Bigl(1-\frac{\beta}{\alpha_1}\Bigr)\nu\bigl[\mathcal{K}(\rho_9,\overline{\pi})\bigr] &\leq \zeta_9\frac{\beta}{\alpha_1}\biggl[\gamma\nu(\rho_9-\overline{\pi})(r)+\mathcal{K}(\nu,\widetilde{\mu}_3) \\
&+\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M')+\alpha_1 R'\bigr\}\Bigr]\Bigr\}+\log(\tfrac{2}{\eta})\biggr] \\
&+\zeta_9\nu\bigl[\mathcal{K}(\rho_9,\pi)-\mathcal{K}(\overline{\pi},\pi)\bigr],
\end{aligned}
$$

where we have introduced a bunch of constants, assumed to be positive, that we will more precisely set to

$$
\begin{aligned}
x_8+x_9&=1, \\
(\zeta_6\beta+x_8\alpha_3)\frac{\gamma}{\alpha_1}&=\lambda_6, \\
(\zeta_7\beta+x_9\alpha_3)\frac{\gamma}{\alpha_1}&=\lambda_7, \\
(\zeta_8\beta-x_8\alpha_3)\frac{\gamma}{\alpha_1}&=\lambda_8,
\end{aligned}
$$

$$(\zeta_9 \beta - x_9 \alpha_3) \frac{\gamma}{\alpha_1} = \lambda_9.$$

We get with $\mathbb{P}$ probability at least $1 - \eta$,

$$\begin{aligned}
\Bigl(1 - \frac{\alpha_3}{\alpha_1} - (\zeta_8 + \zeta_9) \frac{\beta}{\alpha_1}\Bigr) \mathcal{K}(\nu, \widetilde{\mu}_3) \leq \\
\frac{\alpha_3}{\alpha_1} \Bigl[\gamma \bigl[\nu (x_8 \rho_8 + x_9 \rho_9)(r) - \widetilde{\mu}_3 (x_8 \rho_6 + x_9 \rho_7)(r) \bigr] \\
+ \frac{\alpha_3}{\alpha_1} \log \Bigl\{ (\widetilde{\mu}_3 \overline{\pi}) \otimes (\widetilde{\mu}_3 \overline{\pi}) \Bigl[\exp \bigl\{ -\gamma \Psi_{\frac{\gamma}{N}} (R', M') + \alpha_1 R' \bigr\} \Bigr] \Bigr\} \\
+ (\zeta_6 + \zeta_7 + \zeta_8 + \zeta_9) \frac{\beta}{\alpha_1} \log \Bigl\{ \widetilde{\mu}_3 \bigl(\overline{\pi}^{\otimes 2} \bigr) \Bigl[\exp \bigl\{ -\gamma \Psi_{\frac{\gamma}{N}} (R', M') + \alpha_1 R' \bigr\} \Bigr] \Bigr\} \\
+ \mathcal{K}(\nu, \mu) - \mathcal{K}(\widetilde{\mu}_3, \mu) + \Bigl(\frac{\alpha_3}{\alpha_1} + (\zeta_6 + \zeta_7 + \zeta_8 + \zeta_9) \frac{\beta}{\alpha_1} \Bigr) \log \bigl(\tfrac{2}{\eta} \bigr).
\end{aligned}$$

Let us choose the constants so that $\lambda_1 = \lambda_7 = \lambda_9$, $\lambda_4 = \lambda_6 = \lambda_8$, $\alpha_3 x_9 \frac{\gamma}{\alpha_1} = \xi_1$ and $\alpha_3 x_8 \frac{\gamma}{\alpha_1} = \xi_4$. This is done by setting

$$\begin{aligned}
x_8 &= \frac{\xi_4}{\xi_1 + \xi_4}, \\
x_9 &= \frac{\xi_1}{\xi_1 + \xi_4}, \\
\alpha_3 &= \tfrac{N}{\gamma} \sinh(\tfrac{\gamma}{N})(\xi_1 + \xi_4), \\
\zeta_6 &= \tfrac{N}{\gamma} \sinh(\tfrac{\gamma}{N}) \frac{(\lambda_4 - \xi_4)}{\beta}, \\
\zeta_7 &= \tfrac{N}{\gamma} \sinh(\tfrac{\gamma}{N}) \frac{(\lambda_1 - \xi_1)}{\beta}, \\
\zeta_8 &= \tfrac{N}{\gamma} \sinh(\tfrac{\gamma}{N}) \frac{(\lambda_4 + \xi_4)}{\beta}, \\
\zeta_9 &= \tfrac{N}{\gamma} \sinh(\tfrac{\gamma}{N}) \frac{(\lambda_1 + \xi_1)}{\beta}.
\end{aligned}$$

The inequality $\lambda_1 > \xi_1$ is always satisfied. The inequality $\lambda_4 > \xi_4$ is required for the above choice of constants, and will be satisfied for a suitable choice of ζ_3 and ζ_4.

Under these assumptions, we obtain with $\mathbb{P}$ probability at least $1 - \eta$

$$\begin{aligned}
\Bigl(1 - \frac{\alpha_3}{\alpha_1} - (\zeta_8 + \zeta_9) \frac{\beta}{\alpha_1}\Bigr) \mathcal{K}(\nu, \widetilde{\mu}_3) \leq (\nu - \widetilde{\mu}_3)(\xi_1 \rho_1 + \xi_4 \rho_4)(r) \\
+ \frac{\alpha_3}{\alpha_1} \log \Bigl\{ (\widetilde{\mu}_3 \overline{\pi}) \otimes (\widetilde{\mu}_3 \overline{\pi}) \Bigl[\exp \bigl\{ -\gamma \Psi_{\frac{\gamma}{N}} (R', M') + \alpha_1 R' \bigr\} \Bigr] \Bigr\} \\
+ (\zeta_6 + \zeta_7 + \zeta_8 + \zeta_9) \frac{\beta}{\alpha_1} \log \Bigl\{ \widetilde{\mu}_3 \bigl(\overline{\pi}^{\otimes 2} \bigr) \Bigl[\exp \bigl\{ -\gamma \Psi_{\frac{\gamma}{N}} (R', M') + \alpha_1 R' \bigr\} \Bigr] \Bigr\} \\
+ \mathcal{K}(\nu, \mu) - \mathcal{K}(\widetilde{\mu}_3, \mu) + \Bigl(\frac{\alpha_3}{\alpha_1} + (\zeta_6 + \zeta_7 + \zeta_8 + \zeta_9) \frac{\beta}{\alpha_1} \Bigr) \log \bigl(\tfrac{2}{\eta} \bigr).
\end{aligned}$$

This proves

PROPOSITION 2.3.4. *The constants being set as explained above, with $\mathbb{P}$ probability at least $1 - \eta$, for any posterior distribution $\nu : \Omega \to \mathcal{M}_+^1(M)$,*

$$\mathcal{K}(\nu,\widetilde{\mu}_3) \leq \Bigl(1 - \frac{\alpha_3}{\alpha_1} - (\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}\Bigr)^{-1}\Bigl[\mathcal{K}(\nu,\nu_3) + \frac{\alpha_3}{\alpha_1}\log\Bigl\{(\widetilde{\mu}_3\overline{\pi})\otimes(\widetilde{\mu}_3\overline{\pi})\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M') + \alpha_1 R'\bigr\}\Bigr]\Bigr\} + (\zeta_6+\zeta_7+\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M') + \alpha_1 R'\bigr\}\Bigr]\Bigr\} + \Bigl(\frac{\alpha_3}{\alpha_1} + (\zeta_6+\zeta_7+\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}\Bigr)\log\bigl(\tfrac{2}{\eta}\bigr)\Bigr].$$

Thus

$$\mathcal{K}(\nu_3\rho_1,\widetilde{\mu}_3\,\widetilde{\pi}_1) \leq \frac{1+\bigl(1-\frac{\lambda_1}{\gamma}\bigr)^{-1}\frac{\lambda_1}{\gamma}}{1-\frac{\alpha_3}{\alpha_1}-(\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}} \times\Bigl[\frac{\alpha_3}{\alpha_1}\log\Bigl\{(\widetilde{\mu}_3\overline{\pi}\otimes(\widetilde{\mu}_3\overline{\pi})\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M') + \alpha_1 R'\bigr\}\Bigr]\Bigr\} + (\zeta_6+\zeta_7+\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}\log\Bigl\{\widetilde{\mu}_3\bigl(\overline{\pi}^{\otimes 2}\bigr)\Bigl[\exp\bigl\{-\gamma\Psi_{\frac{\gamma}{N}}(R',M') + \alpha_1 R'\bigr\}\Bigr]\Bigr\} + \Bigl(\frac{\alpha_3}{\alpha_1} + (\zeta_6+\zeta_7+\zeta_8+\zeta_9)\frac{\beta}{\alpha_1}\Bigr)\log\bigl(\tfrac{2}{\eta}\bigr)\Bigr] + \Bigl(1-\frac{\lambda_1}{\gamma}\Bigr)^{-1}\frac{\lambda_1}{\gamma}\Bigl[\log\Bigl\{\widetilde{\mu}_3\bigl(\widetilde{\pi}_1^{\otimes 2}\bigr)\Bigl[\exp\bigl\{2N\sinh\bigl(\tfrac{\gamma}{2N}\bigr)^2 M'\bigr\}\Bigr]\Bigr\} - \log\bigl(\tfrac{2}{\eta}\bigr)\Bigr].$$

We will not go further, lest it may become tedious, but we hope we have given sufficient hints to state informally that the bound $B(\nu,\rho,\beta)$ of Theorem 2.3.2 (page 93) is upper bounded with $\mathbb{P}$ probability close to one by a bound of the same flavour where the empirical quantities r and m' have been replaced with their expectations R and M'.

2.3.3. Two step localization between posterior distributions.

Here we work with a family of prior distributions described by a regular conditional prior distribution $\pi = M \to \mathcal{M}_+^1(\Theta)$, where M is some measurable index set. This family may typically describe a countable family of parametric models. In this case $M = \mathbb{N}$, and each of the prior distributions $\pi(i,.)$, $i \in \mathbb{N}$ satisfies some parametric complexity assumption of the type

$$\limsup_{\beta\to+\infty} \beta\bigl[\pi_{\exp(-\beta R)}(i,.)(R) - \operatorname*{ess\,inf}_{\pi(i,.)} R\bigr] = d_i < +\infty, \qquad i \in M.$$

Let us consider also a prior distribution $\mu \in \mathcal{M}_+^1(M)$ defined on the index set M.

Our aim here will be to compare the performance of two given posterior distributions $\nu_1\rho_1$ and $\nu_2\rho_2$, where $\nu_1,\nu_2 : \Omega \to \mathcal{M}_+^1(M)$, and where $\rho_1,\rho_2 : \Omega\times M \to \mathcal{M}_+^1(\Theta)$. More precisely, we would like to establish a bound for $(\nu_1\rho_1 - \nu_2\rho_2)(R)$ which could be a starting point to implement a selection method similar to the one described in Theorem 2.2.4 (page 72). To this purpose, we can start with Theorem 2.2.1 (page 68), which says that with $\mathbb{P}$ probability at least $1-\epsilon$,

$$-N\log\Bigl\{1-\tanh\bigl(\tfrac{\lambda}{N}\bigr)\bigl(\nu_1\rho_1-\nu_2\rho_2\bigr)(R)\Bigr\} \leq \lambda(\nu_1\rho_1-\nu_2\rho_2)(r) + N\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr](\nu_1\rho_1)\otimes(\nu_2\rho_2)(m') + \mathcal{K}(\nu_1,\widetilde{\mu}) + \mathcal{K}(\nu_2,\widetilde{\mu}) + \nu_1\bigl[\mathcal{K}(\rho_1,\widetilde{\pi})\bigr] + \nu_2\bigl[\mathcal{K}(\rho_2,\widetilde{\pi})\bigr] - \log(\epsilon),$$

where $\widetilde{\mu} \in \mathcal{M}_+^1(M)$ and $\widetilde{\pi} : M \rightarrow \mathcal{M}_+^1(\Theta)$ are suitably localized prior distributions to be chosen later on. To use these localized prior distributions, we need empirical bounds for the entropy terms $\mathcal{K}(\nu_i, \widetilde{\mu})$ and $\nu_i\bigl[\mathcal{K}(\rho_i, \widetilde{\pi})\bigr]$, $i = 1, 2$.

Bounding $\nu\bigl[\mathcal{K}(\rho, \widetilde{\pi})\bigr]$ can be done using the following generalization of Corollary 2.1.19 page 68:

Corollary 2.3.5. *For any positive real constants γ and λ such that $\gamma < \lambda$, for any prior distribution $\overline{\mu} \in \mathcal{M}_+^1(M)$ and any conditional prior distribution $\pi : M \rightarrow \mathcal{M}_+^1(\Theta)$, with $\mathbb{P}$ probability at least $1 - \epsilon$, for any posterior distribution $\nu : \Omega \rightarrow \mathcal{M}_+^1(M)$, and any conditional posterior distribution $\rho : \Omega \times M \rightarrow \mathcal{M}_+^1(\Theta)$,*

$$\nu\Bigl\{\mathcal{K}\bigl[\rho, \pi_{\exp[-N\frac{\gamma}{\lambda}\tanh(\frac{\lambda}{N})R]}\bigr]\Bigr\} \leq K'(\nu, \rho, \gamma, \lambda, \epsilon) + \frac{1}{\frac{\lambda}{\gamma} - 1}\mathcal{K}(\nu, \overline{\mu}),$$

where

$$K'(\nu, \rho, \gamma, \lambda, \epsilon) \overset{\text{def}}{=} \Bigl(1 - \tfrac{\gamma}{\lambda}\Bigr)^{-1}\Bigl\{\nu\bigl[\mathcal{K}(\rho, \pi_{\exp(-\gamma r)}\bigr] \\ - \frac{\gamma}{\lambda}\log(\epsilon) + \nu\Bigl\{\log\Bigl[\pi_{\exp(-\gamma r)}\Bigl(\exp\bigl\{N\tfrac{\gamma}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho(m')\bigr\}\Bigr)\Bigr]\Bigr\}\Bigr\}.$$

To apply this corollary to our case, we have to set

$$\widetilde{\pi} = \pi_{\exp[-N\frac{\gamma}{\lambda}\tanh(\frac{\lambda}{N})R]}.$$

Let us also consider for some positive real constant β the conditional prior distribution

$$\overline{\pi} = \pi_{\exp(-\beta R)}$$

and the prior distribution

$$\overline{\mu} = \mu_{\exp[-\alpha\overline{\pi}(R)]}.$$

Let us see how we can bound, given any posterior distribution $\nu : \Omega \rightarrow \mathcal{M}_+^1(M)$, the divergence $\mathcal{K}(\nu, \overline{\mu})$. We can see that

$$\mathcal{K}(\nu, \overline{\mu}) = \alpha(\nu - \overline{\mu})\overline{\pi}(R) + \mathcal{K}(\nu, \mu) - \mathcal{K}(\overline{\mu}, \mu).$$

Now, let us introduce the conditional posterior distribution

$$\widehat{\pi} = \pi_{\exp(-\gamma r)}$$

and let us decompose

$$(\nu - \overline{\mu})\bigl[\overline{\pi}(R)\bigr] = \nu\bigl[\overline{\pi}(R) - \widehat{\pi}(R)\bigr] + (\nu - \overline{\mu})\bigl[\widehat{\pi}(R)\bigr] + \overline{\mu}\bigl[\widehat{\pi}(R) - \overline{\pi}(R)\bigr].$$

Starting from the exponential inequality

$$\mathbb{P}\Bigl[\overline{\mu}\bigl[\overline{\pi} \otimes \overline{\pi}\bigr]\exp\Bigl\{-N\log\bigl[1 - \tanh(\tfrac{\gamma}{N})R'\bigr] - \gamma r' - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\Bigr\}\Bigr] \leq 1,$$

and reasoning in the same way that led to Theorem 2.1.1 (page 52) in the simple case when we take in this theorem $\lambda = \gamma$, we get with $\mathbb{P}$ probability at least $1 - \epsilon$, that

$$- N \log\bigl\{1 - \tanh(\tfrac{\gamma}{N}) \nu(\overline{\pi} - \widehat{\pi})(R)\bigr\} + \beta \nu(\overline{\pi} - \widehat{\pi})(R) \\ \leq \nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \log\bigl[\cosh(\tfrac{\gamma}{N}) \widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] + \mathcal{K}(\nu, \overline{\mu}) - \log(\epsilon).$$

$$- N \log\bigl\{1 - \tanh(\tfrac{\gamma}{N}) \overline{\mu}(\widehat{\pi} - \overline{\pi})(R)\bigr\} - \beta \overline{\mu}(\widehat{\pi} - \overline{\pi})(R) \\ \leq \overline{\mu}\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \log\bigl[\cosh(\tfrac{\gamma}{N}) \widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] - \log(\epsilon).$$

In the meantime, using Theorem 2.2.1 (page 68) and Corollary 2.3.5 above, we see that with $\mathbb{P}$ probability at least $1-2\epsilon$, for any conditional posterior distribution $\rho : \Omega \times M \to \mathcal{M}_+^1(\Theta)$,

$$- N \log\Bigl\{1 - \tanh(\tfrac{\lambda}{N})(\nu - \overline{\mu})\rho(R)\Bigr\} \leq \lambda(\nu - \overline{\mu})\rho(r) \\ + N \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr](\nu\rho) \otimes (\overline{\mu}\rho)(m') + (\nu + \overline{\mu})\mathcal{K}(\rho, \widetilde{\pi}) + \mathcal{K}(\nu, \overline{\mu}) - \log(\epsilon) \\ \leq \lambda(\nu - \overline{\mu})\rho(r) + N \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr](\nu\rho) \otimes (\overline{\mu}\rho)(m') + \mathcal{K}(\nu, \overline{\mu}) - \log(\epsilon) \\ + \Bigl(1 - \tfrac{\gamma}{\lambda}\Bigr)^{-1}(\nu + \overline{\mu})\Bigl\{\mathcal{K}(\rho, \widehat{\pi}) + \log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \tfrac{\gamma}{\lambda} \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\rho(m')\bigr\}\Bigr]\Bigr\}\Bigr\} \\ + \Bigl(\tfrac{\lambda}{\gamma} - 1\Bigr)^{-1}\bigl[\mathcal{K}(\nu, \overline{\mu}) - 2\log(\epsilon)\bigr].$$

Putting all this together, we see that with $\mathbb{P}$ probability at least $1 - 3\epsilon$, for any posterior distribution $\nu \in \mathcal{M}_+^1(M)$,

$$\biggl[1 - \frac{\alpha}{N \tanh(\frac{\gamma}{N}) + \beta} - \frac{\alpha}{N \tanh(\frac{\lambda}{N})\bigl(1 - \frac{\gamma}{\lambda}\bigr)}\biggr]\mathcal{K}(\nu, \overline{\mu}) \leq \\ \alpha\Bigl[N \tanh(\tfrac{\gamma}{N}) + \beta\Bigr]^{-1}\Bigl\{\nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] - \log(\epsilon)\Bigr\} \\ + \alpha\Bigl[N \tanh(\tfrac{\gamma}{N}) - \beta\Bigr]^{-1}\Bigl\{\overline{\mu}\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] - \log(\epsilon)\Bigr\} \\ + \alpha\bigl[N \tanh(\tfrac{\lambda}{N})\bigr]^{-1}\Bigl\{ \\ \lambda(\nu - \overline{\mu})\widehat{\pi}(r) + N \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr](\nu\widehat{\pi}) \otimes (\overline{\mu}\widehat{\pi})(m') \\ + \Bigl(1 - \tfrac{\gamma}{\lambda}\Bigr)^{-1}(\nu + \overline{\mu})\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N \tfrac{\gamma}{\lambda} \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] \\ - \frac{1 + \frac{\gamma}{\lambda}}{1 - \frac{\gamma}{\lambda}} \log(\epsilon)\Bigr\} + \mathcal{K}(\nu, \mu) - \mathcal{K}(\overline{\mu}, \mu).$$

Replacing in the right-hand side of this inequality the unobserved prior distribution $\overline{\mu}$ with the worst possible posterior distribution, we obtain

Theorem 2.3.6. *For any positive real constants α, β, γ and λ, using the notation,*

$$\overline{\pi} = \pi_{\exp(-\beta R)}, \\ \overline{\mu} = \mu_{\exp[-\alpha\overline{\pi}(R)]}, \\ \widehat{\pi} = \pi_{\exp(-\gamma r)},$$

$$\widehat{\mu} = \mu_{\exp[-\alpha \frac{\lambda}{N} \tanh(\frac{\lambda}{N})^{-1} \widehat{\pi}(r)]},$$

with $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\nu : \Omega \rightarrow \mathcal{M}_+^1(M)$,

$$\begin{aligned}
&\biggl[1 - \frac{\alpha}{N\tanh(\frac{\gamma}{N}) + \beta} - \frac{\alpha}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}\biggr] \mathcal{K}(\nu, \overline{\mu}) \leq \mathcal{K}(\nu, \widehat{\mu}) \\
&\quad + \frac{\alpha}{N\tanh(\frac{\gamma}{N}) + \beta}\Bigl\{\nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr]\Bigr\} \\
&\quad + \frac{\alpha}{N\tanh(\frac{\lambda}{N})(1-\frac{\gamma}{\lambda})}\Bigl\{\nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{N\tfrac{\gamma}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr]\Bigr\} \\
&\quad + \log\biggl\{\widehat{\mu}\biggl[\Bigl[\widehat{\pi}\Bigl\{\exp\Bigl[N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\Bigr]\Bigr\}\Bigr]^{\frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}} \\
&\qquad \times \Bigl[\widehat{\pi}\Bigl\{\exp\Bigl[N\tfrac{\gamma}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\widehat{\pi}(m')\Bigr]\Bigr\}\Bigr]^{\frac{\alpha}{N\tanh(\frac{\lambda}{N})(1-\frac{\gamma}{\lambda})}} \\
&\qquad \times \exp\biggl[\frac{\alpha\log[\cosh(\frac{\lambda}{N})]}{\tanh(\frac{\lambda}{N})}(\nu\widehat{\pi})\otimes\widehat{\pi}(m')\biggr]\biggr]\biggr\} \\
&\quad + \biggl[\frac{1}{N\tanh(\frac{\gamma}{N}) + \beta} + \frac{1}{N\tanh(\frac{\gamma}{N}) - \beta} + \frac{1+\frac{\gamma}{\lambda}}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}\biggr]\log\bigl(\tfrac{3}{\epsilon}\bigr).
\end{aligned}$$

This result is satisfactory, but in the same time hints at some possible improvement in the choice of the localized prior $\overline{\mu}$, which is here somewhat lacking a variance term. We will consider in the remainder of this section the use of

$$\overline{\mu} = \mu_{\exp[-\alpha\overline{\pi}(R) - \xi\widetilde{\pi}\otimes\widetilde{\pi}(M')]}, \tag{2.38}$$

where ξ is some positive real constant and $\widetilde{\pi} = \pi_{\exp(-\widetilde{\beta}R)}$ is some appropriate conditional prior distribution with positive real parameter $\widetilde{\beta}$. With this new choice

$$\mathcal{K}(\nu, \overline{\mu}) = \alpha(\nu - \overline{\mu})\overline{\pi}(R) + \xi(\nu - \overline{\mu})(\widetilde{\pi}\otimes\widetilde{\pi})(M') + \mathcal{K}(\nu, \mu) - \mathcal{K}(\overline{\mu}, \mu).$$

We already know how to deal with the first factor $\alpha(\nu - \overline{\mu})\overline{\pi}(R)$, since the computations we made to give it an empirical upper bound were valid for any choice of the localized prior distribution $\overline{\mu}$. Let us now deal with $\xi(\nu - \overline{\mu})(\widetilde{\pi}\otimes\widetilde{\pi})(M')$. Since $m'(\theta, \theta')$ is a sum of independent Bernoulli random variables, we can easily generalize the result of Theorem 1.1.4 (page 4) to prove that with $\mathbb{P}$ probability at least $1-\epsilon$

$$\begin{aligned}
&N\bigl[1 - \exp(-\tfrac{\zeta}{N})\bigr]\nu(\widetilde{\pi}\otimes\widetilde{\pi})(M') \\
&\qquad \leq \zeta\Phi_{\frac{\zeta}{N}}\bigl[\nu(\widetilde{\pi}\otimes\widetilde{\pi})(M')\bigr] \leq \zeta\nu(\widetilde{\pi}\otimes\widetilde{\pi})(m') + \mathcal{K}(\nu, \overline{\mu}) - \log(\epsilon).
\end{aligned}$$

In the same way, with $\mathbb{P}$ probability at least $1-\epsilon$,

$$\begin{aligned}
&-N\bigl[\exp(\tfrac{\zeta}{N}) - 1\bigr]\overline{\mu}(\widetilde{\pi}\otimes\widetilde{\pi})(M') \\
&\qquad \leq -\zeta\Phi_{-\frac{\zeta}{N}}\bigl[\overline{\mu}(\widetilde{\pi}\otimes\widetilde{\pi})(M')\bigr] \leq -\zeta\overline{\mu}(\widetilde{\pi}\otimes\widetilde{\pi})(m') - \log(\epsilon).
\end{aligned}$$

We would like now to replace $(\widetilde{\pi}\otimes\widetilde{\pi})(m')$ with an empirical quantity. In order to do this, we will use an entropy bound. Indeed for any conditional posterior distribution $\rho : \Omega \times M \rightarrow \mathcal{M}_+^1(\Theta)$,

$$\nu\bigl[\mathcal{K}(\rho,\widetilde{\pi})\bigr] = \widetilde{\beta}\nu(\rho-\widetilde{\pi})(R) + \nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\widetilde{\pi},\pi)\bigr]$$
$$\leq \frac{\widetilde{\beta}}{N\tanh(\frac{\gamma}{N})}\biggl\{\gamma\nu(\rho-\widetilde{\pi})(r) + N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\rho\otimes\widetilde{\pi})(m')$$
$$+\mathcal{K}(\nu,\overline{\mu}) + \nu\bigl[\mathcal{K}(\rho,\widetilde{\pi})\bigr] - \log(\epsilon)\biggr\} + \nu\bigl[\mathcal{K}(\rho,\pi)-\mathcal{K}(\widetilde{\pi},\pi)\bigr].$$

Thus choosing $\widetilde{\beta} = N\tanh(\frac{\gamma}{N})$,

$$\gamma\nu(\widetilde{\pi}-\rho)(r) + \nu\bigl[\mathcal{K}(\widetilde{\pi},\pi)-\mathcal{K}(\rho,\pi)\bigr]$$
$$\leq N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\rho\otimes\widetilde{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon).$$

Choosing $\rho = \widehat{\pi}$, we get

$$\nu\bigl[\mathcal{K}(\widetilde{\pi},\widehat{\pi})\bigr] \leq N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\widehat{\pi}\otimes\widetilde{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon).$$

This implies that

$$\xi\nu(\widehat{\pi}\otimes\widetilde{\pi})(m') = \nu\Bigl\{\widetilde{\pi}\bigl[\xi\widehat{\pi}(m')\bigr] - \mathcal{K}(\widetilde{\pi},\widehat{\pi})\Bigr\} + \nu\bigl[\mathcal{K}(\widetilde{\pi},\widehat{\pi})\bigr]$$
$$\leq \nu\Bigl\{\log\Bigl[\widetilde{\pi}\bigl\{\exp\bigl[\xi\widehat{\pi}(m')\bigr]\bigr\}\Bigr]\Bigr\}$$
$$+ N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\nu(\widehat{\pi}\otimes\widetilde{\pi})(m') + \mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon).$$

Thus

$$\bigl\{\xi - N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\bigr\}\nu(\widehat{\pi}\otimes\widetilde{\pi})(m')$$
$$\leq \nu\Bigl\{\log\Bigl[\widetilde{\pi}\bigl\{\exp\bigl[\xi\widehat{\pi}(m')\bigr]\bigr\}\Bigr]\Bigr\} + \mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)$$

and

$$\nu\bigl[\mathcal{K}(\widetilde{\pi},\widehat{\pi})\bigr] \leq \left(\frac{\xi}{N\log[\cosh(\frac{\gamma}{N})]} - 1\right)^{-1}\biggl[\nu\Bigl\{\log\Bigl[\widetilde{\pi}\bigl\{\exp\bigl[\xi\widehat{\pi}(m')\bigr]\bigr\}\Bigr]\Bigr\}$$
$$+\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\biggr] + \mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon).$$

Taking for simplicity $\xi = 2N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]$ and noticing that

$$2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr] = -N\log\bigl(1-\tfrac{\widetilde{\beta}^2}{N^2}\bigr),$$

we get

Theorem 2.3.7. *Let us put* $\widetilde{\pi} = \pi_{\exp(-\widetilde{\beta}R)}$ *and* $\widehat{\pi} = \pi_{\exp(-\gamma r)}$, *where* γ *is some arbitrary positive real constant and* $\widetilde{\beta} = N\tanh(\frac{\gamma}{N})$, *so that* $\gamma = \frac{N}{2}\log\Bigl(\frac{1+\frac{\widetilde{\beta}}{N}}{1-\frac{\widetilde{\beta}}{N}}\Bigr)$. *With* $\mathbb{P}$ *probability at least* $1-\epsilon$,

$$\nu\bigl[\mathcal{K}(\widetilde{\pi},\widehat{\pi})\bigr] \leq \nu\biggl[\log\Bigl\{\widetilde{\pi}\Bigl[\exp\bigl\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\biggr] + 2\bigl[\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\bigr].$$

As a consequence

$$\begin{aligned}\zeta\nu(\widetilde{\pi}\otimes\widetilde{\pi})(m') &= \zeta\nu(\widetilde{\pi}\otimes\widetilde{\pi})(m') - \nu\bigl[\mathcal{K}(\widetilde{\pi}\otimes\widetilde{\pi},\widehat{\pi}\otimes\widehat{\pi})\bigr] + 2\nu\bigl[\mathcal{K}(\widetilde{\pi},\widehat{\pi})\bigr] \\ &\leq \nu\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(\zeta m')\bigr]\Bigr]\Bigr\} \\ &\quad + 2\nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] + 4\bigl[\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\bigr].\end{aligned}$$

Let us take for the sake of simplicity $\zeta = 2N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]$, to get

$$\zeta\nu(\widetilde{\pi}\otimes\widetilde{\pi})(m') \leq 3\nu\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(\zeta m')\bigr]\Bigr]\Bigr\} + 4\bigl[\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\bigr].$$

This proves

Proposition 2.3.8. *Let us consider some arbitrary prior distribution $\overline{\mu} \in \mathcal{M}_+^1(M)$ and some arbitrary conditional prior distribution $\pi : M \to \mathcal{M}_+^1(\Theta)$. Let $\widetilde{\beta} < N$ be some positive real constant. Let us put $\widetilde{\pi} = \pi_{\exp(-\widetilde{\beta}R)}$ and $\widehat{\pi} = \pi_{\exp(-\gamma r)}$, with $\widetilde{\beta} = N\tanh(\frac{\gamma}{N})$. Moreover let us put $\zeta = 2N\log\bigl[\cosh(\frac{\gamma}{N})\bigr]$. With $\mathbb{P}$ probability at least $1-2\epsilon$, for any posterior distribution $\nu \in \mathcal{M}_+^1(M)$,*

$$\begin{aligned}\nu(\widetilde{\pi}\otimes\widetilde{\pi})(M') &\leq \frac{3\nu\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(\zeta m')\bigr]\Bigr]\Bigr\} + 5\bigl[\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\bigr]}{N\bigl[1-\exp(-\frac{\zeta}{N})\bigr]} \\ &= \frac{1}{N\tanh(\frac{\gamma}{N})^2}\Biggl\{3\nu\Bigl[\log\Bigl\{\widehat{\pi}\otimes\widehat{\pi}\Bigl[\exp\bigl\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\bigr\}\Bigr]\Bigr\}\Bigr] \\ &\qquad + 5\bigl[\mathcal{K}(\nu,\overline{\mu}) - \log(\epsilon)\bigr]\Biggr\}.\end{aligned}$$

In the same way,

$$\begin{aligned}-\zeta\overline{\mu}(\widetilde{\pi}\otimes\widetilde{\pi})(m') &\leq \overline{\mu}\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(-\zeta m')\bigr]\Bigr]\Bigr\} \\ &\quad + 2\overline{\mu}\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] - 4\log(\epsilon)\end{aligned}$$

and thus

$$\begin{aligned}-\overline{\mu}(\widetilde{\pi}\otimes\widetilde{\pi})(M') &\leq \frac{1}{N\bigl[\exp(\frac{\zeta}{N})-1\bigr]}\Biggl\{\overline{\mu}\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(-\zeta m')\bigr]\Bigr]\Bigr\} \\ &\quad + 2\overline{\mu}\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\bigl\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\bigr\}\Bigr]\Bigr\}\Bigr] - 5\log(\epsilon)\Biggr\}.\end{aligned}$$

Here we have purposely kept ζ as an arbitrary positive real constant, to be tuned later (in order to be able to strengthen more or less the compensation of variance terms).

We are now properly equipped to estimate the divergence with respect to $\overline{\mu}$, the choice of prior distribution made in equation (2.38, page 104). Indeed we can now write

$$\Biggl[1 - \frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta} - \frac{\alpha}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)} - \frac{5\xi}{N\tanh(\frac{\gamma}{N})^2}\Biggr]\mathcal{K}(\nu,\overline{\mu})$$

$$\leq \frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}\Big\{\nu\Big[\log\Big\{\widehat{\pi}\Big[\exp\big\{N\log\big[\cosh(\tfrac{\gamma}{N})\big]\widehat{\pi}(m')\big\}\Big]\Big\}\Big]-\log(\epsilon)\Big\}$$
$$+\frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}\Big\{\overline{\mu}\Big[\log\Big\{\widehat{\pi}\Big[\exp\big\{N\log\big[\cosh(\tfrac{\gamma}{N})\big]\widehat{\pi}(m')\big\}\Big]\Big\}\Big]-\log(\epsilon)\Big\}$$
$$+\frac{\alpha}{N\tanh(\frac{\lambda}{N})}\Big\{$$
$$\lambda(\nu-\overline{\mu})\widehat{\pi}(r)+N\log\big[\cosh(\tfrac{\lambda}{N})\big](\nu\widehat{\pi})\otimes(\overline{\mu}\widehat{\pi})(m')$$
$$+\Big(1-\tfrac{\gamma}{\lambda}\Big)^{-1}(\nu+\overline{\mu})\Big[\log\Big\{\widehat{\pi}\Big[\exp\big\{N\tfrac{\gamma}{\lambda}\log\big[\cosh(\tfrac{\lambda}{N})\big]\widehat{\pi}(m')\big\}\Big]\Big\}\Big]$$
$$-\frac{1+\frac{\gamma}{N}}{1-\frac{\gamma}{N}}\log(\epsilon)\Big\}$$
$$+\frac{\xi}{N\tanh(\frac{\gamma}{N})^2}\Big\{3\nu\Big[\log\Big\{\widehat{\pi}\otimes\widehat{\pi}\Big[\exp\big\{2N\log\big[\cosh(\tfrac{\gamma}{N})\big]m'\big\}\Big]\Big\}\Big]-5\log(\epsilon)\Big\}$$
$$+\frac{\xi}{N\big[\exp(\frac{\zeta}{N})-1\big]}\Big\{\overline{\mu}\Big\{\log\Big[\widehat{\pi}\otimes\widehat{\pi}\big[\exp(-\zeta m')\big]\Big]\Big\}$$
$$+2\overline{\mu}\Big[\log\Big\{\widehat{\pi}\Big[\exp\big\{2N\log\big[\cosh(\tfrac{\gamma}{N})\big]\widehat{\pi}(m')\big\}\Big]\Big\}\Big]-5\log(\epsilon)\Big\}.$$
$$+\mathcal{K}(\nu,\mu)-\mathcal{K}(\overline{\mu},\mu).$$

It remains now only to replace in the right-hand side of this inequality $\overline{\mu}$ with the worst possible posterior distribution to obtain

THEOREM 2.3.9. *Let $\lambda>\gamma>\beta$, ζ, α and ξ be arbitrary positive real constants. Let us use the notation $\overline{\pi}=\pi_{\exp(-\beta R)}$, $\widetilde{\pi}=\pi_{\exp(-N\tanh(\frac{\gamma}{N})R)}$, $\widehat{\pi}=\pi_{\exp(-\gamma r)}$, $\overline{\mu}=\mu_{\exp[-\alpha\overline{\pi}(R)-\xi\widetilde{\pi}\otimes\widetilde{\pi}(M')]}$ and let us define the posterior distribution $\widehat{\mu}:\Omega\to\mathcal{M}_+^1(M)$ by*

$$\frac{d\widehat{\mu}}{d\mu}\sim\exp\Big\{-\frac{\alpha\lambda}{N\tanh(\frac{\lambda}{N})}\widehat{\pi}(r)+\frac{\xi}{N\big[\exp(\frac{\zeta}{N})-1\big]}\log\Big\{\widehat{\pi}\otimes\widehat{\pi}\big[\exp(-\zeta m')\big]\Big\}\Big\}.$$

Let us assume moreover that

$$\frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}+\frac{\alpha}{N\tanh(\frac{\lambda}{N})(1-\frac{\gamma}{\lambda})}+\frac{5\xi}{N\tanh(\frac{\gamma}{N})^2}<1.$$

With $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\nu:\Omega\to\mathcal{M}_+^1(M)$,

$$\mathcal{K}(\nu,\overline{\mu})\leq\Big[1-\frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}-\frac{\alpha}{N\tanh(\frac{\lambda}{N})\big(1-\frac{\gamma}{\lambda}\big)}-\frac{5\xi}{N\tanh(\frac{\gamma}{N})^2}\Big]^{-1}\Big\{\mathcal{K}(\nu,\widehat{\mu})$$
$$+\frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}\Big\{\nu\Big[\log\Big\{\widehat{\pi}\Big[\exp\big\{N\log\big[\cosh(\tfrac{\gamma}{N})\big]\widehat{\pi}(m')\big\}\Big]\Big\}\Big]\Big\}$$

$$
\begin{aligned}
&+ \frac{\alpha}{N\tanh(\frac{\lambda}{N})(1-\frac{\gamma}{\lambda})}\Bigl\{\nu\Bigl[\log\Bigl\{\widehat{\pi}\Bigl[\exp\{N\tfrac{\gamma}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\widehat{\pi}(m')\}\Bigr]\Bigr\}\Bigr]\Bigr\}\\
&\quad + \frac{\xi}{N\tanh(\frac{\gamma}{N})^2}\Bigl\{3\nu\Bigl[\log\Bigl\{\widehat{\pi}\otimes\widehat{\pi}\Bigl[\exp\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]m'\}\Bigr]\Bigr\}\Bigr]\Bigr\}\\
&\qquad + \frac{\xi}{N\bigl[\exp(\frac{\zeta}{N})-1\bigr]}\Bigl\{\nu\Bigl\{\log\Bigl[\widehat{\pi}\otimes\widehat{\pi}\bigl[\exp(-\zeta m')\bigr]\Bigr]\Bigr\}\Bigr\}\\
&\qquad + \log\Bigl\{\widehat{\mu}\Bigl[\Bigl\{\widehat{\pi}\Bigl[\exp\{N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\}\Bigr]\Bigr\}^{\frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}}\\
&\qquad\quad \times\Bigl\{\widehat{\pi}\Bigl[\exp\{N\tfrac{\gamma}{\lambda}\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\widehat{\pi}(m')\}\Bigr]\Bigr\}^{\frac{\alpha}{N\tanh(\frac{\lambda}{N})\left(1-\frac{\gamma}{\lambda}\right)}}\\
&\qquad\quad \times\Bigl\{\widehat{\pi}\Bigl[\exp\{2N\log\bigl[\cosh(\tfrac{\gamma}{N})\bigr]\widehat{\pi}(m')\}\Bigr]\Bigr\}^{\frac{2\xi}{N\left[\exp(\frac{\zeta}{N})-1\right]}}\\
&\qquad\quad \times\exp\Bigl\{N\log\bigl[\cosh(\tfrac{\lambda}{N})\bigr]\bigl[(\nu\widehat{\pi})\otimes\widehat{\pi}\bigr](m')\Bigr\}\Bigr]\Bigr\}\\
&\quad + \Biggl[\frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}+\frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}+\frac{2\alpha\bigl(1+\frac{\gamma}{N}\bigr)}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}\\
&\qquad\qquad + \frac{5\xi}{N\tanh(\frac{\gamma}{N})^2}+\frac{5\xi}{N\bigl[\exp(\frac{\zeta}{N})-1\bigr]}\Biggr]\log\bigl(\tfrac{5}{\epsilon}\bigr)\Biggr\}.
\end{aligned}
$$

The interest of this theorem lies in the presence of a variance term in the localized posterior distribution $\widehat{\mu}$, which with a suitable choice of parameters seems to be an interesting option in the case when there are nested models: in this situation there may be a need to prevent integration with respect to $\widehat{\mu}$ in the right-hand side to put weight on wild oversized models with large variance terms. Moreover, the right-hand side being empirical, parameters can be, as usual, optimized from data using a union bound on a grid of candidate values.

If one is only interested in the general shape of the result, a simplified inequality as the one below may suffice:

Corollary 2.3.10. *For any positive real constants $\lambda > \gamma > \beta$, ζ, α and ξ, let us use the same notation as in Theorem 2.3.9 (page 107). Let us put moreover*

$$
\begin{aligned}
A_1 &= \frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}+\frac{\alpha}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}+\frac{5\xi}{N\tanh(\frac{\gamma}{N})^2},\\
A_2 &= \frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}+\frac{\alpha}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}+\frac{3\xi}{N\tanh(\frac{\gamma}{N})^2}\\
A_3 &= \frac{\xi}{N\bigl[\exp(\frac{\zeta}{N})-1\bigr]}\\
A_4 &= \frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}+\frac{\alpha}{N\tanh(\frac{\lambda}{N})(1-\frac{\gamma}{\lambda})}+\frac{2\xi}{N[\exp(\frac{\zeta}{N})-1]},\\
A_5 &= \frac{\alpha}{N\tanh(\frac{\gamma}{N})+\beta}+\frac{\alpha}{N\tanh(\frac{\gamma}{N})-\beta}+\frac{2\alpha\bigl(1+\frac{\gamma}{N}\bigr)}{N\tanh(\frac{\lambda}{N})\bigl(1-\frac{\gamma}{\lambda}\bigr)}\\
&\quad + \frac{5\xi}{N\tanh(\frac{\gamma}{N})^2}+\frac{5\xi}{N\bigl[\exp(\frac{\zeta}{N})-1\bigr]},\\
C_1 &= 2N\log\bigl[\cosh\bigl(\tfrac{\lambda}{N}\bigr)\bigr],
\end{aligned}
$$

$$C_2 = N \log\bigl[\cosh(\tfrac{\lambda}{N})\bigr].$$

Let us assume that $A_1 < 1$. With $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distribution $\nu : \Omega \rightarrow \mathcal{M}_+^1(M)$,

$$\begin{aligned}\mathcal{K}(\nu, \overline{\mu}) \leq K(\nu, \alpha, \beta, \gamma, \lambda, \xi, \zeta, \epsilon) \overset{\text{def}}{=} \bigl(1 - A_1\bigr)^{-1} \Biggl\{ \mathcal{K}(\nu, \widehat{\mu}) \\ + A_2 \nu \Bigl[\log \Bigl(\widehat{\pi} \otimes \widehat{\pi} \bigl[\exp \bigl(C_1 m' \bigr) \bigr] \Bigr) \Bigr] + A_3 \nu \Bigl[\log \Bigl(\widehat{\pi} \otimes \widehat{\pi} \bigl[\exp \bigl(-\zeta m' \bigr) \bigr] \Bigr) \Bigr] \\ + \log \biggl\{ \widehat{\mu} \biggl[\Bigl[\widehat{\pi} \Bigl(\exp \bigl[C_1 \widehat{\pi}(m') \bigr] \Bigr) \Bigr]^{A_4} \exp \Bigl(C_2 \bigl[(\nu \widehat{\pi}) \otimes \widehat{\pi} \bigr] (m') \Bigr) \biggr] \biggr\} \\ + A_5 \log \bigl(\tfrac{5}{\epsilon} \bigr) \Biggr\}.\end{aligned}$$

Putting this corollary together with Corollary 2.3.5 (page 102), we obtain

THEOREM 2.3.11. *Let us consider the notation introduced in Corollary 2.3.5 (page 102) and in Theorem 2.3.9 (page 107) and its Corollary 2.3.10 (page 108). Let us consider real positive parameters λ, $\gamma_1' < \lambda_1'$ and $\gamma_2' < \lambda_2'$. Let us consider also two sets of parameters $\alpha_i, \beta_i, \gamma_i, \lambda_i, \xi_i, \zeta_i$, where $i = 1, 2$, both satisfying the conditions stated in Corollary 2.3.10 (page 108). With $\mathbb{P}$ probability at least $1-\epsilon$, for any posterior distributions $\nu_1, \nu_2 : \Omega \rightarrow \mathcal{M}_+^1(M)$, any conditional posterior distributions $\rho_1, \rho_2 : \Omega \times M \rightarrow \mathcal{M}_+^1(\Theta)$,*

$$\begin{aligned}- N \log \Bigl\{ 1 - \tanh \bigl(\tfrac{\lambda}{N} \bigr) \bigl(\nu_1 \rho_1 - \nu_2 \rho_2 \bigr)(R) \Bigr\} \leq \lambda \bigl(\nu_1 \rho_1 - \nu_2 \rho_2 \bigr)(r) \\ + N \log \bigl[\cosh \bigl(\tfrac{\lambda}{N} \bigr) \bigr] \bigl(\nu_1 \rho_1 \bigr) \otimes \bigl(\nu_2 \rho_2 \bigr) \bigl(m' \bigr) \\ + K' \bigl(\nu_1, \rho_1, \gamma_1', \lambda_1', \tfrac{\epsilon}{5} \bigr) + K' \bigl(\nu_2, \rho_2, \gamma_2', \lambda_2', \tfrac{\epsilon}{5} \bigr) \\ + \frac{1}{1 - \frac{\gamma_1'}{\lambda_1'}} K \bigl(\nu_1, \alpha_1, \beta_1, \gamma_1, \lambda_1, \xi_1, \zeta_1, \tfrac{\epsilon}{5} \bigr) \\ + \frac{1}{1 - \frac{\gamma_2'}{\lambda_2'}} K \bigl(\nu_2, \alpha_2, \beta_2, \gamma_2, \lambda_2, \xi_2, \zeta_2, \tfrac{\epsilon}{5} \bigr) - \log \bigl(\tfrac{\epsilon}{5} \bigr).\end{aligned}$$

This theorem provides, using a union bound argument to further optimize the parameters, an empirical bound for $\nu_1 \rho_1(R) - \nu_2 \rho_2(R)$, which can serve to build a selection algorithm exactly in the same way as what was done in Theorem 2.2.4 (page 72). This represents the highest degree of sophistication that we will achieve in this monograph, as far as model selection is concerned: this theorem shows that it is indeed possible to derive a selection scheme in which localization is performed in two steps and in which the localization of the model selection itself, as opposed to the localization of the estimation in each model, includes a variance term as well as a bias term, so that it should be possible to localize the choice of nested models, something that would not have been feasible with the localization techniques exposed in the previous sections of this study. We should point out however that *more sophisticated* does not necessarily mean *more efficient*: as the reader may have noticed, sophistication comes at a price, in terms of the complexity of the estimation schemes, with some possible loss of accuracy in the constants that can mar the benefits of using an asymptotically more efficient method for small sample sizes.

We will do the hurried reader a favour: we will not launch into a study of the theoretical properties of this selection algorithm, although it is clear that all the tools needed are at hand!

We would like as a conclusion to this chapter, to put forward a simple idea: this approach of model selection revolves around entropy estimates concerned with the divergence of posterior distributions with respect to localized prior distributions. Moreover, this localization of the prior distribution is more effectively done in several steps in some situations, and it is worth mentioning that these situations include the typical case of selection from a family of parametric models. Finally, the whole story relies upon estimating the relative generalization error rate of one posterior distribution with respect to some local prior distribution as well as with respect to another posterior distribution, because these relative rates can be estimated more accurately than absolute generalization error rates, at least as soon as no classification model of reasonable size provides a good match to the training sample, meaning that the classification problem is either difficult or noisy.