Weierstraß-Institut für
Angewandte Analysis und Stochastik

Leibniz
Gemeinschaft

Bayesian inference and stochastic optimization

Vladimir Spokoiny ,
WIAS, HU Berlin

26.09.2019

Outline

Let $\boldsymbol{Y}$ denote the observed random data, $\boldsymbol{Y} \sim I\!\!P$. A general parametric approach (PA): $I\!\!P \in \bigl(I\!\!P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta \subseteq I\!\!R^{\infty}\bigr)$.

The log-likelihood function

$$L(\boldsymbol{\theta}) = L(\boldsymbol{Y}, \boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{dI\!\!P_{\boldsymbol{\theta}}}{d\boldsymbol{\mu}_0}(\boldsymbol{Y}).$$

MLE of $\boldsymbol{\theta}$ by maximizing the function $L(\boldsymbol{\theta})$:

$$\widetilde{\boldsymbol{\theta}} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}).$$

Target (the best parametric fit):

$$\boldsymbol{\theta}^{*} \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} I\!\!E L(\boldsymbol{\theta}).$$

In the Bayes setup $\boldsymbol{\vartheta}$ is a random element, $\boldsymbol{\vartheta} \sim \mathit{\Pi}$ on the parameter set $\mathit{\Theta}$, a prior.

The posterior describes the conditional distribution of $\boldsymbol{\vartheta}$ given $\boldsymbol{Y}$

$$\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y} \propto \exp\bigl\{L(\boldsymbol{\theta})\bigr\}\, \mathit{\Pi}(d\boldsymbol{\theta}).$$

A Gaussian prior $\mathit{\Pi} \sim \mathcal{N}(0, G^{-2})$ yields

$$\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y} \propto \exp\bigl\{L_G(\boldsymbol{\theta})\bigr\},$$

$$L_G(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - \bigl\|G\boldsymbol{\theta}\bigr\|^2/2.$$

Prior $\mathit{\Pi}$ induces a penalty $-\log \mathit{\Pi}(\boldsymbol{\theta})$ leading to penalized MLE

$$\widetilde{\boldsymbol{\theta}}_{\mathit{\Pi}} = \operatorname*{argmax}_{\boldsymbol{\theta}} L_{\mathit{\Pi}}(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \bigl\{ L_{\mathit{\Pi}}(\boldsymbol{\theta}) - \log \mathit{\Pi}(\boldsymbol{\theta}) \bigr\}.$$

Gaussian priors lead to quadratic penalization

$$\widetilde{\boldsymbol{\theta}}_{G} = \operatorname*{argmax}_{\boldsymbol{\theta}} L_{G}(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \bigl\{ L_{G}(\boldsymbol{\theta}) - \bigl\| G\boldsymbol{\theta} \bigr\|^{2}/2 \bigr\}.$$

Posterior $\boldsymbol{\vartheta}_{G} \bigm| \boldsymbol{Y}$ is a random measure with the density

$$\boldsymbol{\vartheta}_{G} \bigm| \boldsymbol{Y} \propto \exp L_{G}(\boldsymbol{\theta})$$

Formally, Bayes approach replaces the max of L_{G} by the soft-max.

- Posterior mean and MAP

$$\widetilde{\boldsymbol{\theta}}_{\mathrm{MAP}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \exp L_G(\boldsymbol{\theta}) = \widetilde{\boldsymbol{\theta}}_G$$

$$\overline{\boldsymbol{\theta}}_G \stackrel{\mathrm{def}}{=} \frac{\int \boldsymbol{\theta} \exp L_G(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int \exp L_G(\boldsymbol{\theta}) d\boldsymbol{\theta}} ;$$

- Concentration (contraction) set $\mathcal{A}$:

$$\rho(\mathcal{A}) \stackrel{\mathrm{def}}{=} \frac{\int_{\mathcal{A}^c} \exp L_G(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\mathcal{A}} \exp L_G(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

- Credible sets $\mathcal{A}(\alpha)$

$$I\!\!P\bigl(\boldsymbol{\vartheta}_G \in \mathcal{A}(\alpha) \bigm| \boldsymbol{Y}\bigr) = 1 - \alpha;$$

- Elliptic credible sets $\mathcal{A}_G(\alpha) = \bigl\{\boldsymbol{\theta} \colon \bigl\| D\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G\bigr)\bigr\| \le z_\alpha \bigr\}$.

$$\widetilde{\boldsymbol{\theta}}_G = \operatorname*{argmax}_{\boldsymbol{\theta}} L_G(\boldsymbol{\theta}), \qquad \boldsymbol{\theta}_G^* = \operatorname*{argmax}_{\boldsymbol{\theta}} \mathbb{E} L_G(\boldsymbol{\theta})$$

- Elliptic concentration sets around $\boldsymbol{\theta}_G^*$:

$$\mathbb{P}\bigl(\bigl\| D_G\bigl(\widetilde{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G^*\bigr)\bigr\| > \mathtt{r}_G(\mathtt{x})\bigr) \le \mathrm{e}^{-\mathtt{x}},$$

where

$$D_G^2 = -\nabla^2 \mathbb{E} L_G(\boldsymbol{\theta}_G^*)\,,$$

$$\mathtt{r}_G(\mathtt{x}) \approx \sqrt{\mathtt{p}_G} + \sqrt{2\mathtt{x}}\,,$$

$$\mathtt{p}_G = \operatorname{tr}\bigl(D_G^{-2}\operatorname{Var}(\nabla L_G(\boldsymbol{\theta}_G^*))\bigr).$$

- Fisher expansion

$$\bigl\| D_G\bigl(\widetilde{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G^*\bigr) - D_G^{-1}\nabla L_G(\boldsymbol{\theta}_G^*)\bigr\| \le 3\mathtt{r}_G^{-1}\delta_{3,G}(\mathtt{r}_G).$$

Contraction sets $\mathcal{A}$: for $\mathtt{x} > 0$ build $\mathcal{A}$ s.t.

$$\rho(\mathcal{A}) \stackrel{\text{def}}{=} \frac{\int_{\mathcal{A}^c} \exp\bigl\{ L_G(\boldsymbol{\theta}) \bigr\} d\boldsymbol{\theta}}{\int_{\mathcal{A}} \exp\bigl\{ L_G(\boldsymbol{\theta}) \bigr\} d\boldsymbol{\theta}} \le \mathrm{e}^{-\mathtt{x}} \quad \text{whp}$$

in the form

$$\mathcal{A} = \widetilde{\mathcal{A}}(\mathtt{r}_0) = \bigl\{ \boldsymbol{\theta} \colon \| \widetilde{D}(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}_G) \| \le \mathtt{r}_0 \bigr\},$$

Gaussian approximation:

$$\sup_{A \in \mathcal{B}_s(\mathbb{R}^p)} \Bigl| \mathbb{P}\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G \in A \,\big|\, \boldsymbol{Y} \bigr) - \mathbb{P}'\bigl(\widetilde{D}_G^{-1} \boldsymbol{\gamma} \in A \bigr) \Bigr| \lesssim \frac{\mathtt{p}_G + \mathtt{x}}{n}$$

$$\sup_{A \in \mathcal{B}(\mathbb{R}^p)} \Bigl| \mathbb{P}\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G \in A \,\big|\, \boldsymbol{Y} \bigr) - \mathbb{P}'\bigl(\widetilde{D}_G^{-1} \boldsymbol{\gamma} \in A \bigr) \Bigr| \lesssim \frac{(\mathtt{p}_G + \mathtt{x})^{3/2}}{n^{1/2}}$$

- Log-likelihood $L(\boldsymbol{\theta}) = L(\boldsymbol{Y}, \boldsymbol{\theta})$.
- Prior $\varPi$ and penalized log-likelihood

$$L_{\varPi}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) + \log \varPi(\boldsymbol{\theta})$$

- Gaussian prior $\mathcal{N}(0, G^{-2})$ and quadratic penalization

$$L_G(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - \|G\boldsymbol{\theta}\|^2/2$$

- Penalized MLE

$$\widetilde{\boldsymbol{\theta}}_G = \operatorname*{argmax}_{\boldsymbol{\theta}} L_G(\boldsymbol{\theta})$$

- Posterior $\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y}$

$$\boldsymbol{\vartheta} \,\big|\, \boldsymbol{Y} \propto \exp L_G(\boldsymbol{\theta})$$

True value

$$\boldsymbol{\theta}^* = \operatorname*{argmax}_{\boldsymbol{\theta}} I\!E L(\boldsymbol{\theta}).$$

- The penalized MLE $\widetilde{\boldsymbol{\theta}}_G$ concentrates around

$$\boldsymbol{\theta}_G^* = \operatorname*{argmax}_{\boldsymbol{\theta}} I\!E L_G(\boldsymbol{\theta})$$

  yielding the bias $\boldsymbol{\theta}_G^* - \boldsymbol{\theta}^*$.
- The posterior measure $\boldsymbol{\vartheta} \bigm| \boldsymbol{Y}$ concentrates around $\widetilde{\boldsymbol{\theta}}_G$
- $\boldsymbol{\vartheta} \bigm| \boldsymbol{Y}$ is nearly Gaussian.

$(\boldsymbol{E})$ *The stochastic component* $\zeta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - I\!\!E L(\boldsymbol{\theta})$ *of the process* $L(\boldsymbol{\theta})$ *is linear in* $\boldsymbol{\theta}$ *:*

$$\nabla\zeta \equiv \nabla\zeta(\boldsymbol{\theta}).$$

$(\boldsymbol{ED_0})$ *There exist a positive symmetric matrix* V *, and constants* $\mathrm{g} > 0$ *,* $\nu_0 \geq 1$ *such that* $\mathrm{Var}\bigl(\nabla\zeta\bigr) \leq V^2$ *and*

$$\sup_{\boldsymbol{u} \in I\!\!R^p} \log I\!\!E \exp\biggl\{\lambda \frac{\langle \boldsymbol{u}, \nabla\zeta \rangle}{\|V\boldsymbol{u}\|}\biggr\} \leq \frac{\nu_0^2 \lambda^2}{2}, \qquad |\lambda| \leq \mathrm{g}.$$

$(\mathcal{L})$ *The set* Θ *is open and convex in* $\mathbb{R}^p$ *. The function* $\mathbb{E}L(\boldsymbol{\theta})$ *is concave in* $\boldsymbol{\theta} \in \Theta$ *.*

Define for each $\boldsymbol{\theta} \in \Theta^{\circ}$, and any $\boldsymbol{u} \in \mathbb{R}^p$, the directional derivative

$$\delta_m(\boldsymbol{\theta}, \boldsymbol{u}) \stackrel{\mathrm{def}}{=} \frac{1}{m!} \frac{d^m}{dt^m} \mathbb{E}L(\boldsymbol{\theta} + t\boldsymbol{u}) \Big|_{t=0}, \quad m = 3, 4.$$

$(\mathcal{L}_{\mathbf{0}})$ *The functions* $\delta_3(\boldsymbol{\theta}, \boldsymbol{u})$ *and* $\delta_4(\boldsymbol{\theta}, \boldsymbol{u})$ *are well defined and uniformly bounded for all* $\boldsymbol{\theta} \in \Theta^{\circ}$ *and all* $\boldsymbol{u} \in \mathcal{U}$ *for specific sets* Θ° *and* $\mathcal{U}$ *in* $\mathbb{R}^p$ *.*

Clearly the value $\delta_3(\boldsymbol{\theta}, \boldsymbol{u})$ is proportional to $\|\boldsymbol{u}\|^3$, while $\delta_4(\boldsymbol{\theta}, \boldsymbol{u}) \asymp \|\boldsymbol{u}\|^4$.

Define

$$D^2(\boldsymbol{\theta}) = -\nabla^2 I\!E L(\boldsymbol{\theta}), \quad D_G^2 = D^2(\boldsymbol{\theta}) + G^2.$$

$(\boldsymbol{V}|\boldsymbol{G})$ Signal-noise:

$$B_{V|G}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} D_G^{-1}(\boldsymbol{\theta})\, V^2\, D_G^{-1}(\boldsymbol{\theta})$$

with V^2 from $(\boldsymbol{ED_0})$. There are fixed constants $\lambda_{V|G}$ and $\mathtt{p}_{V|G}$ such that

$$\operatorname{tr} B_{V|G}(\boldsymbol{\theta}) \le \mathtt{p}_{V|G}\,, \qquad \|B_{V|G}\| \le \lambda_{V|G}\,, \qquad \boldsymbol{\theta} \in \Theta^{\circ}\,.$$

$(\boldsymbol{D}|\boldsymbol{G})$ The local effective dimension: for a fixed constant $\mathtt{C}$

$$\mathtt{p}_G(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \operatorname{tr}\bigl\{ D^2(\boldsymbol{\theta}) D_G^{-2}(\boldsymbol{\theta}) \bigr\} \le \mathtt{C}, \qquad \boldsymbol{\theta} \in \Theta^{\circ}.$$

By $(\boldsymbol{E})$, the stochastic component $\zeta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - I\!E L(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$ and $\nabla\zeta = \nabla\zeta(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\theta}$.

Theorem

Under condition $(\boldsymbol{ED_0})$, there exists a random set $\Omega(\mathtt{x})$ with $I\!P\bigl(\Omega(\mathtt{x})\bigr) \geq 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$ such that on this set

$$\left\| D_G^{-1} \nabla\zeta \right\| \leq z(B_{V|G}, \mathtt{x}),$$

where $B_{V|G} = D_G^{-1} V^2 D_G^{-1}$ and

$$z(B_{V|G}, \mathtt{x}) = \sqrt{\operatorname{tr} B_{V|G}} + \sqrt{2\mathtt{x}\lambda_{\max}(B_{V|G})}.$$

Let $\boldsymbol{\theta}_G^* = \operatorname{arginf}_{\boldsymbol{\theta}} \mathbb{E} L_G(\boldsymbol{\theta})$ and $D_G^2 = D^2(\boldsymbol{\theta}_G^*) + G^2$.

Theorem

Let $\bigl\| D_G^{-1} \nabla \zeta \bigr\| \le z(B_{V|G}, \mathtt{x})$ *on a random set* $\Omega(\mathtt{x})$ *with* $\mathbb{P}\bigl(\Omega(\mathtt{x})\bigr) \ge 1 - \mathrm{e}^{-\mathtt{x}}$. *Let*

$$\mathcal{A}_G(\mathtt{r}_G) \stackrel{\text{def}}{=} \bigl\{ \boldsymbol{\theta} \colon \| D_G(\boldsymbol{\theta} - \boldsymbol{\theta}_G^*) \| \le \mathtt{r}_G \bigr\}$$

$$(1-\rho)\mathtt{r}_G \ \ge\ z(B_{V|G}, \mathtt{x}).$$

Then on $\Omega(\mathtt{x})$

$$\bigl\| D_G \bigl(\widetilde{\boldsymbol{\theta}}_G - \boldsymbol{\theta}_G^* \bigr) \bigr\| \le \mathtt{r}_G .$$

Local set

$$\mathcal{A}_G(\mathtt{r}_G) \stackrel{\text{def}}{=} \bigl\{\boldsymbol{\theta} \colon \|D_G(\boldsymbol{\theta} - \boldsymbol{\theta}_G^*)\| \le \mathtt{r}_G\bigr\}.$$

Use local smoothness of $-I\!\!E L_G(\boldsymbol{\theta})$ to show that for $\boldsymbol{\theta} \in \mathcal{A}_G(\mathtt{r}_G)$

$$-\bigl\{I\!\!E L_G(\boldsymbol{\theta}) - I\!\!E L_G(\boldsymbol{\theta}_G^*)\bigr\} \approx \bigl\|D_G(\boldsymbol{\theta} - \boldsymbol{\theta}_G^*)\bigr\|^2/2,$$

$$-\nabla I\!\!E L_G(\boldsymbol{\theta}) \approx D_G^2(\boldsymbol{\theta} - \boldsymbol{\theta}_G^*)$$

Use convexity of $I\!\!E L_G(\boldsymbol{\theta})$ to show that

$$\bigl\|\nabla I\!\!E L_G(\boldsymbol{\theta})\bigr\| \ge \mathtt{r}_G\,, \quad \boldsymbol{\theta} \notin \mathcal{A}_G(\mathtt{r}_G).$$

Use $\bigl\|D_G^{-1}\nabla\zeta\bigr\| \le z(B_{V|G}, \mathtt{x})$ to show

$$\nabla L(\boldsymbol{\theta}) = \nabla I\!\!E L(\boldsymbol{\theta}) + \nabla\zeta \ne 0, \quad \boldsymbol{\theta} \notin \mathcal{A}_G(\mathtt{r}_G).$$

Concentration sets of the posterior in nonparametric models using

- Empirical process theory, large deviations of the log-likelihood and covering numbers and chaining arguments
- small ball probability
- local smoothness

See e.g.

- Ghoshal, S., Ghosh, J. K., van der Vaart, A. W. (2000) Convergence rates of posterior distributions. Ann.Statist., 28, 500–531.
- van der Vaart, A. W., van Zanten, J. H. (2008) Rates of contraction of posterior distributions based on Gaussian process priors. Ann. Statist., 36, 1031–1508.

Define the elliptic set $\widetilde{\mathcal{A}}(\mathtt{r}_0) = \bigl\{\boldsymbol{u}\colon \|\widetilde{D}\boldsymbol{u}\| \le \mathtt{r}_0\bigr\}$ with $\widetilde{D}^2 = D^2(\widetilde{\boldsymbol{\theta}}_G)$. Consider the random quantity

$$\rho(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \frac{\int_{\widetilde{\mathcal{A}}(\mathtt{r}_0)^c} \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u})\bigr\} d\boldsymbol{u}}{\int_{\widetilde{\mathcal{A}}(\mathtt{r}_0)} \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u})\bigr\} d\boldsymbol{u}} .$$

Theorem

Let, for some fixed values $\mathtt{r}_0$ and $\mathtt{x} > 0$, it hold

$$\mathtt{C}_0 \mathtt{r}_0 \ge 2\sqrt{\mathtt{p}_G(\boldsymbol{\theta})} + \sqrt{\mathtt{x}}\,, \qquad \boldsymbol{\theta} \in \varTheta^{\circ} .$$

Then, on the random set $\varOmega(\mathtt{x})$ from Theorem 2, with $\widetilde{\mathtt{p}}_G = \mathtt{p}_G(\widetilde{\boldsymbol{\theta}}_G)$

$$\rho(\mathtt{r}_0) \le \exp\bigl\{-(\widetilde{\mathtt{p}}_G + \mathtt{x})/2\bigr\}.$$

Let $\widetilde{\boldsymbol{\theta}}_G = \operatorname{arginf}_{\boldsymbol{\theta}} L_G(\boldsymbol{\theta})$. The use of $\nabla L_G(\widetilde{\boldsymbol{\theta}}_G) = 0$ allows to represent

$$\rho(\mathtt{r}_0) = \frac{\int_{\|\widetilde{D}\boldsymbol{u}\| > \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) \bigr\} d\boldsymbol{u}}{\int_{\|\widetilde{D}\boldsymbol{u}\| \le \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) \bigr\} d\boldsymbol{u}}$$

$$= \frac{\int_{\|\widetilde{D}\boldsymbol{u}\| > \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) - \bigl\langle \nabla L_G(\widetilde{\boldsymbol{\theta}}_G), \boldsymbol{u} \bigr\rangle \bigr\} d\boldsymbol{u}}{\int_{\|\widetilde{D}\boldsymbol{u}\| \le \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) - \bigl\langle \nabla L_G(\widetilde{\boldsymbol{\theta}}_G), \boldsymbol{u} \bigr\rangle \bigr\} d\boldsymbol{u}} .$$

Fix $\boldsymbol{\theta} \in \mathcal{A}_G(\mathtt{r}_G)$. Consider $f(\boldsymbol{\theta}) = \mathbb{E} L_G(\boldsymbol{\theta})$. As $\zeta(\boldsymbol{\theta}) = L(\boldsymbol{\theta}) - \mathbb{E} L(\boldsymbol{\theta})$ is linear in $\boldsymbol{\theta}$, it holds

$$
\begin{aligned}
& L_G(\boldsymbol{\theta} + \boldsymbol{u}) - L_G(\boldsymbol{\theta}) - \bigl\langle \nabla L_G(\boldsymbol{\theta}), \boldsymbol{u} \bigr\rangle \\
& \qquad = f(\boldsymbol{\theta} + \boldsymbol{u}) - f(\boldsymbol{u}) - \bigl\langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \bigr\rangle .
\end{aligned}
$$

Therefore, it suffices to bound the ratio

$$
\rho(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{\int_{\mathcal{U}^c} \exp\bigl\{ f(\boldsymbol{\theta} + \boldsymbol{u}) - f(\boldsymbol{u}) - \bigl\langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \bigr\rangle \bigr\} d\boldsymbol{u}}{\int_{\mathcal{U}} \exp\bigl\{ f(\boldsymbol{\theta} + \boldsymbol{u}) - f(\boldsymbol{u}) - \bigl\langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \bigr\rangle \bigr\} d\boldsymbol{u}}
$$

for the elliptic set $\mathcal{U} = \mathcal{U}(\boldsymbol{\theta}, \mathtt{r}_0) = \bigl\{ \boldsymbol{u} \colon \| D(\boldsymbol{\theta}) \boldsymbol{u} \| \le \mathtt{r}_0 \bigr\}$ uniformly in $\boldsymbol{\theta}$ from the set $\bigl\{ \boldsymbol{\theta} \colon \bigl\| D_G \bigl(\boldsymbol{\theta} - \boldsymbol{\theta}^*_G \bigr) \bigr\| \le \mathtt{r}_G \bigr\}$; see Theorem 2.

First we present some bounds for the denominator of $\rho(\boldsymbol{\theta})$. Local smoothness of $f(\boldsymbol{\theta}) = I\!E L_G(\boldsymbol{\theta})$ implies

$$\begin{aligned}\int_{\mathcal{U}} & \exp\bigl\{f(\boldsymbol{\theta}+\boldsymbol{u}) - f(\boldsymbol{u}) - \bigl\langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u}\bigr\rangle\bigr\}\, d\boldsymbol{u} \\ &\approx \int_{\mathcal{U}} \exp\Bigl(-\frac{\|D_G(\boldsymbol{\theta})\boldsymbol{u}\|^2}{2}\Bigr)\, d\boldsymbol{u},\end{aligned}$$

Moreover,

$$\frac{\det D_G(\boldsymbol{\theta})}{(2\pi)^{p/2}} \int_{\mathcal{U}} \exp\Bigl(-\frac{\|D_G(\boldsymbol{\theta})\boldsymbol{u}\|^2}{2}\Bigr)\, d\boldsymbol{u} = I\!P\Bigl(\bigl\|D(\boldsymbol{\theta})D_G^{-1}(\boldsymbol{\theta})\boldsymbol{\gamma}\bigr\| \le \mathtt{r}_0\Bigr)$$

for $\boldsymbol{\gamma} \sim \mathcal{N}(0, I_p)$. The choice $\mathtt{r}_0 \ge \sqrt{\mathtt{p}_G(\boldsymbol{\theta})} + \sqrt{2\mathtt{x}}$ yields

$$I\!P\Bigl(\bigl\|D(\boldsymbol{\theta})D_G^{-1}(\boldsymbol{\theta})\boldsymbol{\gamma}\bigr\| \le \mathtt{r}_0\Bigr) \ge 1 - \mathrm{e}^{-\mathtt{x}}.$$

Step 3: $\int_{\mathcal{U}^c} \exp\{ f(\boldsymbol{\theta}+\boldsymbol{u}) - f(\boldsymbol{u}) - \langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \rangle \} d\boldsymbol{u}$

$f(\boldsymbol{\theta}) = \mathbb{E} L(\boldsymbol{\theta})$ is concave and $-\langle \nabla^2 f(\boldsymbol{\theta}) \boldsymbol{u}, \boldsymbol{u} \rangle = \| D(\boldsymbol{\theta}) \boldsymbol{u} \|^2$.
For any $\boldsymbol{u}$ with $\| D(\boldsymbol{\theta}) \boldsymbol{u} \| = \mathtt{r} > \mathtt{r}_0$

$$
\begin{aligned}
& f(\boldsymbol{\theta}+\boldsymbol{u}) - f(\boldsymbol{\theta}) - \langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \rangle - \| G \boldsymbol{u} \|^2/2 \\
& \quad \leq -\mathtt{C}_0 (\| D(\boldsymbol{\theta}) \boldsymbol{u} \| \mathtt{r}_0 - \mathtt{r}_0^2/2) - \| G \boldsymbol{u} \|^2/2 \\
& \quad = -\mathtt{C}_0 (\| D(\boldsymbol{\theta}) \boldsymbol{u} \| \mathtt{r}_0 - \mathtt{r}_0^2/2) - \| D_G(\boldsymbol{\theta}) \boldsymbol{u} \|^2/2 + \| D(\boldsymbol{\theta}) \boldsymbol{u} \|^2/2.
\end{aligned}
$$

with $\mathtt{C}_0 \geq 1/2$ and $D_G^2(\boldsymbol{\theta}) = D^2(\boldsymbol{\theta}) + G^2$.
Now we can use the result about Gaussian integrals:

$$
\begin{aligned}
& \frac{\det D_G}{(2\pi)^{p/2}} \int_{\| D \boldsymbol{u} \| \geq \mathtt{r}_0} \exp\bigl\{ - \bigl(\| D \boldsymbol{u} \| \mathtt{r}_0 - \mathtt{r}_0^2/2 - \| D_G \boldsymbol{u} \|^2 + \| D \boldsymbol{u} \|^2 \bigr)/2 \bigr\} \, d\boldsymbol{u} \\
& \quad \leq \mathtt{C} \mathrm{e}^{-(\mathtt{p}_G(\boldsymbol{\theta}) + \mathtt{x})/2}.
\end{aligned}
$$

Outline

- van der Vaart, A. W. , van Zanten, J. H. (2008)
- Leahu, H. (2011). Gaussian models
- Castillo, I. , Nickl, R. (2013, 2014). General results.
- Panov and Sp (2015). Semiparametric problems

Contraction results and local quadratic approximation of the log-likelihood. The Gaussian approximation in a weak sense.

Theorem

Suppose that It holds on $\Omega(\mathtt{x})$ *with* $\widetilde{D}_G^2 = D_G^2(\widetilde{\boldsymbol{\theta}}_G)$

$$\sup_{A \in \mathcal{B}_s(\mathbb{R}^p)} \Bigl| \mathbb{P}\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - \mathbb{P}'\bigl(\widetilde{D}_G^{-1}\boldsymbol{\gamma} \in A\bigr) \Bigr| \le \mathtt{C}\diamondsuit(\mathtt{r}_0)$$

$$\sup_{A \in \mathcal{B}(\mathbb{R}^p)} \Bigl| \mathbb{P}\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - \mathbb{P}'\bigl(\widetilde{D}_G^{-1}\boldsymbol{\gamma} \in A\bigr) \Bigr| \le \mathtt{C}\delta_3(\mathtt{r}_0)$$

with

$$\delta_3(\mathtt{r}_0) \lesssim \frac{\mathtt{r}_0^3}{\sqrt{n}} \lesssim \frac{\mathtt{p}_G^{3/2}}{\sqrt{n}},$$

$$\diamondsuit(\mathtt{r}_0) \lesssim \frac{\mathtt{r}_0^6}{n} \lesssim \frac{\mathtt{p}_G^3}{n}.$$

Fix any centrally symmetric set A. First we restrict the posterior probability to the set $\widetilde{\mathcal{A}}(\mathtt{r}_0) = \{\boldsymbol{u}\colon \|\widetilde{D}\boldsymbol{u}\| \le \mathtt{r}_0\}$. Then we apply the quadratic approximation of the log-likelihood function $L(\boldsymbol{\theta})$. Denote $A(\mathtt{r}_0) = A \cap \widetilde{\mathcal{A}}(\mathtt{r}_0)$. Obviously, $A(\mathtt{r}_0)$ is centrally symmetric as well. Further,

$$\mathbb{P}\bigl(\boldsymbol{\vartheta}_G - \widetilde{\boldsymbol{\theta}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) = \frac{\int_A \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u})\bigr\} d\boldsymbol{u}}{\int_{\mathbb{R}^p} \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u})\bigr\} d\boldsymbol{u}}$$

$$\le \frac{\int_{A(\mathtt{r}_0)} \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) - \bigl\langle \nabla L_G(\widetilde{\boldsymbol{\theta}}_G), \boldsymbol{u}\bigr\rangle\bigr\} d\boldsymbol{u}}{\int_{\widetilde{\mathcal{A}}(\mathtt{r}_0)} \exp\bigl\{L_G(\widetilde{\boldsymbol{\theta}}_G + \boldsymbol{u}) - L_G(\widetilde{\boldsymbol{\theta}}_G) - \bigl\langle \nabla L_G(\widetilde{\boldsymbol{\theta}}_G), \boldsymbol{u}\bigr\rangle\bigr\} d\boldsymbol{u}} + \rho(\mathtt{r}_0).$$

Fix $\boldsymbol{\theta} \in \mathcal{A}_G(\mathtt{r}_G)$. Then for $A \subset \mathcal{U}$ centrally symmetric

$$\int_A \exp\bigl\{ f(\boldsymbol{\theta} + \boldsymbol{u}) - f(\boldsymbol{u}) - \langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \rangle \bigr\} \, d\boldsymbol{u}$$

$$\geq \bigl(1 - \diamondsuit(\mathtt{r}_0)\bigr) \int_A \exp\Bigl(- \frac{\| D_G(\boldsymbol{\theta}) \boldsymbol{u} \|^2}{2} \Bigr) \, d\boldsymbol{u},$$

$$\int_A \exp\bigl\{ f(\boldsymbol{\theta} + \boldsymbol{u}) - f(\boldsymbol{u}) - \langle \nabla f(\boldsymbol{\theta}), \boldsymbol{u} \rangle \bigr\} \, d\boldsymbol{u}$$

$$\leq \bigl(1 + \diamondsuit(\mathtt{r}_0)\bigr) \int_A \exp\Bigl(- \frac{\| D_G(\boldsymbol{\theta}) \boldsymbol{u} \|^2}{2} \Bigr) \, d\boldsymbol{u},$$

where $\diamondsuit(\mathtt{r}_0) = 4\delta_3^2 + 4\delta_4$.

Consider the model

$$\boldsymbol{Y} = A(\boldsymbol{f}^*) + \sigma\boldsymbol{\varepsilon} \in \mathcal{Y}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I) \tag{1}$$

with a known non-linear operator $A\colon \mathcal{X} \to \mathcal{Y}$ for Hilbert spaces $\mathcal{X}, \mathcal{Y}$.
We assume that A is injective and denote by $\mathfrak{A}$ its inverse:
$\mathfrak{A} \stackrel{\text{def}}{=} A^{-1}\colon \mathcal{Y} \to \mathcal{X}$.
A Gaussian log-likelihood in the model (1) w.r.t. $\boldsymbol{f}$ is of the form

$$L(\boldsymbol{f}) = -(2\sigma^2)^{-1}\big\|\boldsymbol{Y} - A(\boldsymbol{f})\big\|^2.$$

A Gaussian prior $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}_0, G^{-2})$ yields the penalized log-likelihood

$$\mathcal{L}_G(\boldsymbol{f}) = -(2\sigma^2)^{-1}\big\|\boldsymbol{Y} - A(\boldsymbol{f})\big\|^2 - \frac{1}{2}\big\|G(\boldsymbol{f} - \boldsymbol{f}_0)\big\|^2.$$

However, it does not satisfy the crucial conditions $(\mathcal{L})$ and $(\boldsymbol{E})$.

One can use another parametrization by $\boldsymbol{g} = A(\boldsymbol{f}) \in \mathcal{Y}$ where $\boldsymbol{g}$ is "smooth". This yields the classical log-likelihood

$$L(\boldsymbol{g}) = -\frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{g}\|^2.$$

Let now $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}_0, G^{-2})$. It leads to a non-Gaussian prior $\mathfrak{A}(\boldsymbol{g})$ on $\mathcal{Y}$. The penalized log-likelihood $\mathcal{L}_G(\boldsymbol{g})$ reads

$$\mathcal{L}_G(\boldsymbol{g}) = -\frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{g}\|^2 - \frac{1}{2}\|G\{\mathfrak{A}(\boldsymbol{g}) - \boldsymbol{f}_0\}\|^2 + \mathtt{C}.$$

Further, it holds for the expected log-likelihood

$$\mathbb{E}\mathcal{L}_G(\boldsymbol{\eta}) = -\frac{1}{2\sigma^2}\|A(\boldsymbol{f}^*) - \boldsymbol{g}\|^2 - \frac{1}{2}\|G\{\mathfrak{A}(\boldsymbol{g}) - \boldsymbol{f}_0\}\|^2 + \mathtt{C}$$

Problem: $\mathfrak{A}$ unbounded and non-smooth.

Represent $\boldsymbol{Y} = A(\boldsymbol{f}) + \sigma\boldsymbol{\varepsilon}$ by two identities $\boldsymbol{Y} = \boldsymbol{g} + \sigma\boldsymbol{\varepsilon}$ and $\boldsymbol{g} = A(\boldsymbol{f})$. Then the fidelity terms can be relaxed to

$$\sigma^{-2}\big\|\boldsymbol{Y} - \boldsymbol{g}\big\|^2 + \lambda\|\boldsymbol{g} - A(\boldsymbol{f})\|^2.$$

Bayesian modeling assumes regular priors on $\boldsymbol{g}$ and $\boldsymbol{f}$. In particular, one can use Gaussian priors $\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{f}_0, G^{-2})$ and $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{g}_0, \varGamma^{-2})$ yielding the posterior

$$\boldsymbol{f}, \boldsymbol{g} \,\big|\, \boldsymbol{Y} \propto \exp \mathcal{L}_G(\boldsymbol{f}, \boldsymbol{g})$$

$$\begin{aligned} \mathcal{L}_G(\boldsymbol{f}, \boldsymbol{g}) = & -\frac{1}{2\sigma^2}\big\|\boldsymbol{Y} - \boldsymbol{g}\big\|^2 - \frac{\lambda}{2}\|\boldsymbol{g} - A(\boldsymbol{f})\|^2 \\ & -\frac{1}{2}\big\langle \boldsymbol{f} - \boldsymbol{f}_0 \big\rangle_G - \frac{1}{2}\big\langle \boldsymbol{g} - \boldsymbol{g}_0 \big\rangle_{\varGamma} . \end{aligned}$$

The gradient and Hessian of $\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g})$ read as follow:

$$\frac{d}{d\boldsymbol{f}}\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g}) = -\lambda\left\{A(\boldsymbol{f})-\boldsymbol{g}\right\}^\top \nabla A(\boldsymbol{f}) - G^2(\boldsymbol{f}-\boldsymbol{f}_0)$$

$$\frac{d}{d\boldsymbol{g}}\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g}) = \sigma^{-2}\big(\boldsymbol{Y}-\boldsymbol{g}\big) + \lambda\left\{A(\boldsymbol{f})-\boldsymbol{g}\right\} - \mathit{\Gamma}^2\big(\boldsymbol{g}-\boldsymbol{g}_0\big)$$

and

$$\begin{aligned}\mathbb{F}(\boldsymbol{f},\boldsymbol{g}) &\stackrel{\text{def}}{=} -\nabla^2 I\!\!E\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g}) \\ &= \begin{pmatrix} G^2+\lambda\,(\nabla^\top\nabla+\boldsymbol{\delta}^\top\nabla^2)A(\boldsymbol{f}) & \lambda\,\nabla A(\boldsymbol{f}) \\ \lambda\left\{\nabla A(\boldsymbol{f})\right\}^\top & \mathit{\Gamma}^2+(\sigma^{-2}+\lambda)I \end{pmatrix}\end{aligned}$$

with the elasticity vector $\boldsymbol{\delta} = A(\boldsymbol{f}) - \boldsymbol{g}$. For convexity of $-I\!\!E\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g})$ it suffices to check that $\nabla^2 I\!\!E\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g}) < 0$.

In a special case when A is linear and $\nabla A(\boldsymbol{f}) = A$, it holds $\nabla^2 A = 0$ and the matrix $\mathbb{F}(\boldsymbol{f}, \boldsymbol{g})$ does not depend on $(\boldsymbol{f}, \boldsymbol{g})$:

$$\mathbb{F} = -\nabla^2 I\!E \mathcal{L}_G(\boldsymbol{f}, \boldsymbol{g}) = \begin{pmatrix} G^2 + \lambda A^\top A & \lambda A \\ \lambda A^\top & \mathit{\Gamma}^2 + (\sigma^{-2} + \lambda) I \end{pmatrix}.$$

With $G^2 > 0$ and $\mathit{\Gamma}^2 > 0$, this matrix is also positive semidefinite. The $\boldsymbol{f}\boldsymbol{f}$-block of the matrix $\mathbb{F}^{-1}$ reads

$$\begin{aligned} (\mathbb{F}^{-1})_{\boldsymbol{f}\boldsymbol{f}} &= \bigl(\mathbb{F}_{\boldsymbol{f}\boldsymbol{f}} - \mathbb{F}_{\boldsymbol{f}\boldsymbol{g}} \mathbb{F}_{\boldsymbol{g}\boldsymbol{g}}^{-1} \mathbb{F}_{\boldsymbol{g}\boldsymbol{f}}\bigr)^{-1} \\ &= \bigl(G^2 + \lambda A^\top A\bigr)^{-1} - \lambda^2 A^\top \bigl\{\mathit{\Gamma}^2 + (\sigma^{-2} + \lambda) I\bigr\}^{-1} A. \end{aligned}$$

One can see that each of regularizations by G^2 and $\mathit{\Gamma}^2$ improves the conditioning number of $\mathbb{F}_{\boldsymbol{f}\boldsymbol{f}} - \mathbb{F}_{\boldsymbol{f}\boldsymbol{g}} \mathbb{F}_{\boldsymbol{g}\boldsymbol{g}}^{-1} \mathbb{F}_{\boldsymbol{g}\boldsymbol{f}}$. Moreover, one can use $\lambda = \sigma^{-2}$ yielding $(\mathbb{F}^{-1})_{\boldsymbol{f}\boldsymbol{f}} \asymp \mathbb{F}_{\boldsymbol{f}\boldsymbol{f}}^{-1} = \bigl\{G^2 + \sigma^{-2} A^\top A\bigr\}^{-1}$.

In the case of a non-linear operator A

$$
\begin{aligned}
\mathbb{F}(\boldsymbol{f}, \boldsymbol{g}) &\stackrel{\text{def}}{=} -\nabla^2 I\!\!E \mathcal{L}_G(\boldsymbol{f}, \boldsymbol{g}) \\
&= \begin{pmatrix} G^2 + \lambda\,(\nabla^\top \nabla + \boldsymbol{\delta}^\top \nabla^2) A(\boldsymbol{f}) & \lambda\,\nabla A(\boldsymbol{f}) \\ \lambda\,\{\nabla A(\boldsymbol{f})\}^\top & \mathit{\Gamma}^2 + (\sigma^{-2} + \lambda) I \end{pmatrix}
\end{aligned}
$$

with the elasticity vector $\boldsymbol{\delta} = A(\boldsymbol{f}) - \boldsymbol{g}$.

The only term in the matrix $\mathbb{F}(\boldsymbol{f}, \boldsymbol{g})$ involving the second derivative of $A(\cdot)$ appears in the $\boldsymbol{f}\boldsymbol{f}$-block with the multiplicative elasticity factor $\boldsymbol{\delta} = A(\boldsymbol{f}) - \boldsymbol{g}$.

If the structural equation $A(\boldsymbol{f}) = \boldsymbol{g}$ is nearly fulfilled, then this factor vanishes and the impact of this nonlinearity disappears.

The partial penalized MLE in $\boldsymbol{g}$ given $\boldsymbol{f}$ can be computed explicitly.
The equation $\frac{d}{d\boldsymbol{g}}\mathcal{L}_G(\boldsymbol{f},\boldsymbol{g}) = 0$ yields

$$\widetilde{\boldsymbol{g}}(\boldsymbol{f}) = \bigl\{(\sigma^{-2}+\lambda)I + \mathit{\Gamma}^2\bigr\}^{-1}\bigl\{\sigma^{-2}\boldsymbol{Y} + \lambda A(\boldsymbol{f}) + \mathit{\Gamma}^2\boldsymbol{g}_0\bigr\}.$$

With $\boldsymbol{\eta}$ fixed, optimization w.r.t. $\boldsymbol{f}$ yields the equation

$$\lambda\bigl\{A(\boldsymbol{f}) - \boldsymbol{g}\bigr\}^{\top}\nabla A(\boldsymbol{f}) = -G^2(\boldsymbol{f} - \boldsymbol{f}_0).$$

With a guess $\boldsymbol{f}^{\circ}$ and $\boldsymbol{\delta} = A(\boldsymbol{f}^{\circ}) - \boldsymbol{g}$, one gets
$A(\boldsymbol{f}) \approx A(\boldsymbol{f}^{\circ}) + \nabla A(\boldsymbol{f}^{\circ})\,(\boldsymbol{f} - \boldsymbol{f}^{\circ})$ and
$\nabla A(\boldsymbol{f}) \approx \nabla A(\boldsymbol{f}^{\circ}) + \nabla^2 A(\boldsymbol{f}^{\circ})\,(\boldsymbol{f} - \boldsymbol{f}^{\circ})$. This yields an update

$$\begin{aligned}\boldsymbol{f} - \boldsymbol{f}^{\circ} = {} & \bigl\{G^2 + \lambda(\nabla^{\top}\nabla + \boldsymbol{\delta}^{\top}\nabla^2)A(\boldsymbol{f}^{\circ})\bigr\}^{-1} \\ & \times\bigl[G^2(\boldsymbol{f}^{\circ} - \boldsymbol{f}_0) + \lambda\,\boldsymbol{\delta}^{\top}\nabla A(\boldsymbol{f}^{\circ})\bigr]\end{aligned}$$

Now we present a result about large deviation bound for the penalized MLE. Let

$$\widetilde{\boldsymbol{\upsilon}}_G = \big(\widetilde{\boldsymbol{f}}_G, \widetilde{\boldsymbol{g}}_G\big) \stackrel{\text{def}}{=} \operatorname*{argmax}_{\boldsymbol{\upsilon} \in \varUpsilon^{\circ}} \mathcal{L}_G(\boldsymbol{\upsilon}).$$

Under condition $(\boldsymbol{ED_0})$, for any $\mathtt{x} > 0$ there exists a random set $\Omega(\mathtt{x})$ with $I\!\!P\big(\Omega(\mathtt{x})\big) \geq 1 - \mathtt{C}\mathrm{e}^{-\mathtt{x}}$ such that $\nabla_{\boldsymbol{g}}\zeta(\boldsymbol{g})$ satisfies

$$\big\| D_G^{-1} \nabla_{\boldsymbol{g}} \zeta \big\| \leq z(B_{V|G}, \mathtt{x}) \text{ on } \Omega(\mathtt{x}),$$

where $B_{V|G} = D_G^{-1} V^2 D_G^{-1}$ and for any matrix B

$$z(B, x) = \sqrt{\operatorname{tr} B} + \sqrt{2\lambda_{\max}(B)\mathtt{x}}.$$

Theorem

Let $\left\| D_G^{-1} \nabla_{\boldsymbol{g}} \zeta \right\| \leq z(B_{V|G}, \mathtt{x})$ *hold on a random set* $\Omega(\mathtt{x})$ *with* $\mathbb{P}\bigl(\Omega(\mathtt{x})\bigr) \geq 1 - \mathrm{e}^{-\mathtt{x}}$ *and*

$$\mathtt{r}_G \geq 2z(B_{V|G}, \mathtt{x}).$$

Then on $\Omega(\mathtt{x})$

$$\left\| \mathcal{D}_G \bigl(\widetilde{\boldsymbol{v}}_G - \boldsymbol{v}_G^{*} \bigr) \right\| \leq \mathtt{r}_G \, .$$

This also implies

$$\left\| D_G \bigl(\widetilde{\boldsymbol{f}}_G - \boldsymbol{f}_G^{*} \bigr) \right\| \leq \mathtt{r}_G \, .$$

Theorem

Let

$$\mathtt{C}_0 \mathtt{r}_0 \geq 2\sqrt{\mathrm{p}_G(\boldsymbol{\upsilon})} + \sqrt{\mathtt{x}}\,, \qquad \boldsymbol{\upsilon} \in \varUpsilon_0 \,.$$

Then, on the random set $\Omega(\mathtt{x})$ *from Theorem 5, the quantity*

$$\rho(\mathtt{r}_0) \stackrel{\mathrm{def}}{=} \frac{\int_{\|\widetilde{\mathbb{D}}\boldsymbol{u}\| \geq \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\upsilon}}_G + \boldsymbol{u}) \bigr\} d\boldsymbol{u}}{\int_{\|\widetilde{\mathbb{D}}\boldsymbol{u}\| \leq \mathtt{r}_0} \exp\bigl\{ L_G(\widetilde{\boldsymbol{\upsilon}}_G + \boldsymbol{u}) \bigr\} d\boldsymbol{u}}$$

fulfills

$$\rho(\mathtt{r}_0) \leq 2 \exp\bigl\{ -(\widetilde{\mathrm{p}}_G + \mathtt{x})/2 \bigr\} \,,$$

where $\widetilde{\mathrm{p}}_G = \mathrm{p}_G(\widetilde{\boldsymbol{\upsilon}}_G)$ *.*

Corollary

It holds on $\Omega(\mathtt{x})$

$$\sup_{A \in \mathcal{B}_s(\varUpsilon^d)} \left| I\!\!P\bigl(\boldsymbol{\upsilon}_G - \widetilde{\boldsymbol{\upsilon}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - I\!\!P'\bigl(\widetilde{\mathbb{D}}_G^{-1}\boldsymbol{\gamma} \in A\bigr) \right| \lesssim \diamondsuit(\mathtt{r}_0)$$

$$\sup_{A \in \mathcal{B}(\varUpsilon^d)} \left| I\!\!P\bigl(\boldsymbol{\upsilon}_G - \widetilde{\boldsymbol{\upsilon}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - I\!\!P'\bigl(\widetilde{\mathbb{D}}_G^{-1}\boldsymbol{\gamma} \in A\bigr) \right| \lesssim \delta_3(\mathtt{r}_0).$$

Finally we state the result about the marginal posterior $\boldsymbol{f}_G \,\big|\, \boldsymbol{Y}$.

Corollary

It holds on $\Omega(\mathtt{x})$

$$\sup_{A \in \mathcal{B}_s(\mathcal{X}^d)} \left| I\!P\bigl(\boldsymbol{f}_G - \widetilde{\boldsymbol{f}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - I\!P'\bigl(\widetilde{D}_G^{-1} \boldsymbol{\gamma} \in A\bigr) \right| \lesssim \diamondsuit(\mathtt{r}_0)$$

$$\sup_{A \in \mathcal{B}(\mathcal{X}^d)} \left| I\!P\bigl(\boldsymbol{f}_G - \widetilde{\boldsymbol{f}}_G \in A \,\big|\, \boldsymbol{Y}\bigr) - I\!P'\bigl(\widetilde{D}_G^{-1} \boldsymbol{\gamma} \in A\bigr) \right| \lesssim \delta_3(\mathtt{r}_0).$$