Weierstraß-Institut für
Angewandte Analysis und Stochastik

Leibniz-Gemeinschaft

Big ball probability with applications

Vladimir Spokoiny, WIAS and HU (Berlin)
joint with Friedrich Götze (Bielefeld), Alexey Naumov (Skoltech, Moscow) and Vladimir Ulyanov (MSU, Moscow)

Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be Gaussian elements in a Hilbert space $\mathcal{H}$ with zero mean and covariance operators $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ respectively. For any $\boldsymbol{a} \in \mathcal{H}$, we are interested in bounding

$$\sup_{x>0} \left| \mathbb{P}(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - \mathbb{P}(\|\boldsymbol{\eta}\| \le x) \right|$$

in terms of $\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}$ and $\|\boldsymbol{a}\|$.

A related problem: to bound

$$\mathbb{P}(x < \|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x + \Delta)$$

for small Δ in terms of $\Sigma_{\boldsymbol{\xi}} = \operatorname{Var}(\boldsymbol{\xi})$ and $\|\boldsymbol{a}\|$.

In this paper

$$\mathbb{P}(x < \|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x + \Delta) \lesssim \mathtt{C}_2 \, \Delta,$$

with an explicit constant $\mathtt{C}_2 = \mathtt{C}_2(\Sigma_{\boldsymbol{\xi}}, \boldsymbol{a})$.

All bounds are "dimension free".

Consider an independent sample $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top \sim I\!P$.
The parametric ML approach: $I\!P = I\!P_{\boldsymbol{\theta}^*} \in \bigl(I\!P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta \subseteq I\!R^p \bigr) \ll \boldsymbol{\mu}$.
The corresponding log-likelihood function $L(\boldsymbol{\theta})$:

$$L(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \frac{dI\!P_{\boldsymbol{\theta}}}{d\boldsymbol{\mu}}(\boldsymbol{Y}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}), \qquad \ell_i(Y_i, \boldsymbol{\theta}) = \log \frac{dP_{i,\boldsymbol{\theta}}}{d\mu_i}(Y_i).$$

The MLE $\widetilde{\boldsymbol{\theta}}$ of the true parameter $\boldsymbol{\theta}^*$ is the maximizer of $L(\boldsymbol{\theta})$:

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

If the parametric assumption is misspecified then

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

LR based confidence set (CS) $\mathcal{E}(\mathfrak{z})$ for $\boldsymbol{\theta}^*$ is given by

$$\mathcal{E}(\mathfrak{z}) \stackrel{\text{def}}{=} \{\boldsymbol{\theta} \colon L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) \le \mathfrak{z}\}.$$

The value $\mathfrak{z}$ should be selected to ensure given α that

$$I\!P\bigl(\boldsymbol{\theta}^* \notin \mathcal{E}(\mathfrak{z})\bigr) = I\!P\bigl(L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^*) > \mathfrak{z}\bigr) \approx \alpha.$$

However, choice of $\mathfrak{z}$ depends on the unknown measure $I\!P$.

Bootstrap: apply with i.i.d. random weights $w_i^\flat$, e.g. $w_i^\flat \sim \mathcal{N}(1,1)$:

$$L^\flat(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}) w_i^\flat .$$

Define the bootstrap quantile $\mathfrak{z}^\flat$ by

$$I\!P^\flat\bigl(\widetilde{\boldsymbol{\theta}} \notin \mathcal{E}^\flat(\mathfrak{z}^\flat)\bigr) = I\!P\Bigl(\sup_{\boldsymbol{\theta}\in\Theta} L^\flat(\boldsymbol{\theta}) - L^\flat(\widetilde{\boldsymbol{\theta}}) > \mathfrak{z}^\flat \,\Big|\, \boldsymbol{Y}\Bigr) = \alpha.$$

Bootstrap consistency means that for n large

$$\mathbb{P}\big(\boldsymbol{\theta}^{*} \notin \mathcal{E}(\mathfrak{z}^{\flat})\big) = \mathbb{P}\big(L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}^{*}) > \mathfrak{z}^{\flat}\big) \approx \alpha;$$

see e.g. [Spokoiny and Zhilova, 2015]. The key steps are two approximations:

$$2L(\widetilde{\boldsymbol{\theta}}) - 2L(\boldsymbol{\theta}^{*}) \overset{\mathsf{d}}{\approx} \big\|\boldsymbol{\xi} + \boldsymbol{a}\big\|^{2}, \qquad 2L^{\flat}(\widetilde{\boldsymbol{\theta}}^{\flat}) - 2L^{\flat}(\widetilde{\boldsymbol{\theta}}) \overset{\mathsf{d}}{\approx} \big\|\boldsymbol{\xi}^{\flat}\big\|^{2} \tag{1}$$

where $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma)$, $\boldsymbol{\xi}^{\flat} \,\big|\, \boldsymbol{Y} \sim \mathcal{N}(0, \Sigma^{\flat})$ with Σ given by

$$\Sigma \overset{\text{def}}{=} D^{-1} \operatorname{Var}\big[\nabla L(\boldsymbol{\theta}^{*})\big] D^{-1}, \qquad D^{2} = -\nabla^{2} \mathbb{E} L(\boldsymbol{\theta}^{*}),$$

$$\Sigma^{\flat} \overset{\text{def}}{=} D^{-1} \left(\sum_{i=1}^{n} \nabla \ell_{i}(Y_{i}, \boldsymbol{\theta}^{*}) \big\{\nabla \ell_{i}(Y_{i}, \boldsymbol{\theta}^{*})\big\}^{\top} \right) D^{-1}.$$

The vector $\boldsymbol{a}$ in (1) is the so called modeling bias. It vanishes if the parametric assumption $\mathbb{P} = \mathbb{P}_{\boldsymbol{\theta}^{*}}$ is precisely fulfilled.

So, the quality of the bootstrap approximation can be measured by the distance between two Gaussian distributions $\mathcal{N}(\boldsymbol{a}, \varSigma)$ and $\mathcal{N}(0, \varSigma^{\flat})$.

The matrix Bernstein inequality implies

$$\|\varSigma^{\flat} - \varSigma\| \lesssim \sqrt{\frac{\log p}{n}}.$$

[Spokoiny and Zhilova, 2015] used the Pinsker inequality:

$$\left\|\mathcal{N}(\boldsymbol{a}, \varSigma) - \mathcal{N}(0, \varSigma^{\flat})\right\|_{\mathrm{TV}} \leq \frac{1}{2}\Bigl(\left\|\varSigma^{-1/2}\varSigma^{\flat}\varSigma^{-1/2} - I_p\right\|_{\mathrm{Fr}} + \left\|\varSigma^{-1/2}\boldsymbol{a}\right\|\Bigr);$$

However, the bounds can be significantly improved if consider CS of the form

$$\mathcal{E}(\mathfrak{z}) = \bigl\{\boldsymbol{\theta} \colon L(\widetilde{\boldsymbol{\theta}}) - L(\boldsymbol{\theta}) \leq \mathfrak{z}\bigr\} \approx \mathbb{I}\bigl\{\|\boldsymbol{\xi}\|^2 \leq 2\mathfrak{z}\bigr\}.$$

Namely, under some technical conditions

$$\left| I\!P\left(\|\boldsymbol{\xi} + \boldsymbol{a}\|^2 > \mathfrak{z}^{\flat} \right) - \alpha \right| \leq \frac{\mathtt{C}}{\|\Sigma\|_{\mathrm{Fr}}} \left(\|\Sigma - \Sigma^{\flat}\|_1 + \|\boldsymbol{a}\|^2 \right),$$

where $\|\Sigma\|_{\mathrm{Fr}}$ is the Frobenius norm of Σ.

In particular, the "small modeling bias" condition " $\|\Sigma^{-1/2}\boldsymbol{a}\|$ is small" from [Spokoiny and Zhilova, 2015] is relaxed to " $\|\boldsymbol{a}\|^2 / \|\Sigma\|_{\mathrm{Fr}}$ is small".

Also, the Frobenius norm $\left\|\Sigma^{-1/2}\Sigma^{\flat}\Sigma^{-1/2} - I_p\right\|_{\mathrm{Fr}}$ can be much larger than the normalized nuclear norm $\left\|\Sigma - \Sigma^{\flat}\right\|_1 / \|\Sigma\|_{\mathrm{Fr}}$ if the eigenvalues of Σ and $\Sigma^{\flat}$ rapidly decrease.

Consider a linear regression model

$$Y_i = \Psi_i^\top \boldsymbol{\theta} + \varepsilon_i$$

For homogeneous Gaussian errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, the log-likelihood reads

$$L(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \Psi_i^\top \boldsymbol{\theta})^2 + R = -\frac{1}{2\sigma^2} \left\| \boldsymbol{Y} - \boldsymbol{\Psi}^\top \boldsymbol{\theta} \right\|^2 + R,$$

A Gaussian prior $\Pi = \Pi_G = \mathcal{N}\left(0, G^{-2}\right)$ results in the posterior

$$\begin{aligned} \boldsymbol{\vartheta}_G \,\big|\, \boldsymbol{Y} &\propto \exp\left(L(\boldsymbol{\theta}) - \frac{1}{2} \|G\boldsymbol{\theta}\|^2 \right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2} \left\| \boldsymbol{Y} - \boldsymbol{\Psi}^\top \boldsymbol{\theta} \right\|^2 - \frac{1}{2} \|G\boldsymbol{\theta}\|^2 \right). \end{aligned}$$

Represent $L_G(\boldsymbol{\theta}) \stackrel{\text{def}}{=} L(\boldsymbol{\theta}) - \frac{1}{2}\|G\boldsymbol{\theta}\|^2$ in the form

$$L_G(\boldsymbol{\theta}) = L_G(\breve{\boldsymbol{\theta}}_G) - \frac{1}{2}\big\|D_G(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G)\big\|^2,$$

where

$$\breve{\boldsymbol{\theta}}_G \stackrel{\text{def}}{=} \big(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2\big)^{-2}\boldsymbol{\Psi}\boldsymbol{Y},$$

$$D_G^2 \stackrel{\text{def}}{=} \sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2.$$

In particular, it implies that the posterior distribution $I\!\!P(\boldsymbol{\vartheta}_G \,\big|\, \boldsymbol{Y})$ of $\boldsymbol{\vartheta}_G$ given $\boldsymbol{Y}$ is $\mathcal{N}(\breve{\boldsymbol{\theta}}_G, D_G^{-2})$.
Posterior contraction property means concentration of the posterior on

$$E_G(\mathtt{r}) = \big\{\boldsymbol{\theta}\colon \|W(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G)\| \le \mathtt{r}\big\},$$

where W is a given linear mapping from $I\!\!R^p$.

An elliptic credible set:

$$E_G(\mathtt{r}) = \{\boldsymbol{\theta} \colon \|W(\boldsymbol{\theta} - \breve{\boldsymbol{\theta}}_G)\| \le \mathtt{r}\},$$

The desirable credibility property manifests the prescribed conditional probability of $\boldsymbol{\vartheta}_G \in E(\mathtt{r}_G)$ given $\boldsymbol{Y}$ with $\mathtt{r}_G$ defined for a given α by

$$I\!P\Bigl(\bigl\|W(\boldsymbol{\vartheta}_G - \breve{\boldsymbol{\theta}}_G)\bigr\| \ge \mathtt{r}_G \,\big|\, \boldsymbol{Y}\Bigr) = \alpha.$$

Under the posterior measure $\boldsymbol{\vartheta}_G \sim \mathcal{N}(\breve{\boldsymbol{\theta}}_G, D_G^{-2}) \,\big|\, \boldsymbol{Y}$, this bound reads as

$$I\!P\bigl(\|\boldsymbol{\xi}_G\| \ge \mathtt{r}_G\bigr) = \alpha$$

with a zero mean normal vector $\boldsymbol{\xi}_G \sim \mathcal{N}(0, \Sigma_G)$ for $\Sigma_G = W D_G^{-2} W^{\top}$.

Prior impact: how the credible probability depends on the prior covariance G.

Consider another prior $\varPi_1 = \mathcal{N}(0, G_1^{-2})$ with the covariance matrix G_1^{-2}.
The corresponding posterior $\boldsymbol{\vartheta}_{G_1}$ is again normal but now with parameters $\breve{\boldsymbol{\theta}}_{G_1} = \bigl(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G_1^2\bigr)^{-2}\boldsymbol{\Psi}\boldsymbol{Y}$ and $D_{G_1}^2 = \sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G_1^2$, and

$$I\!\!P\Bigl(\bigl\|W\bigl(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\| \ge \mathtt{r}_G \,\big|\, \boldsymbol{Y}\Bigr) = I\!\!P\Bigl(\bigl\|\boldsymbol{\xi}_{G_1} - \boldsymbol{a}\bigr\| \ge \mathtt{r}_G\Bigr)$$

with $\boldsymbol{\xi}_{G_1} \sim \mathcal{N}(0, \varSigma_{G_1})$ for $\varSigma_{G_1} = W D_{G_1}^{-2} W^\top$ and

$$\boldsymbol{a} \stackrel{\text{def}}{=} W\bigl(\breve{\boldsymbol{\theta}}_G - \breve{\boldsymbol{\theta}}_{G_1}\bigr).$$

Therefore,

$$\Bigl|I\!\!P\Bigl(\bigl\|W\bigl(\boldsymbol{\vartheta}_{G_1} - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\| \ge \mathtt{r}_G \,\big|\, \boldsymbol{Y}\Bigr) - \alpha\Bigr|$$
$$\le \sup_{\mathtt{r}>0}\Bigl|I\!\!P\Bigl(\bigl\|\boldsymbol{\xi}_{G_1} - \boldsymbol{a}\bigr\| \ge \mathtt{r}\Bigr) - I\!\!P\Bigl(\bigl\|\boldsymbol{\xi}_G\bigr\| \ge \mathtt{r}\Bigr)\Bigr| .$$

Consider the Gaussian Bayes model for $\boldsymbol{Y} \in I\!R^n$, $\boldsymbol{\theta} \in I\!R^p$,

$$\boldsymbol{Y} \,\big|\, \boldsymbol{\theta} \propto \exp L(\boldsymbol{\theta}), \quad L(\boldsymbol{\theta}) = \|\boldsymbol{Y} - \Psi^{\top}\boldsymbol{\theta}\|^2/(2\sigma^2),$$

$$\boldsymbol{\vartheta} \propto \mathcal{N}(0, G^{-2})$$

Nonparametric Bayes: the true DGP is $\boldsymbol{Y} = \boldsymbol{\Psi}^{\top}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$.

Fix the radius $\mathtt{r} = \mathtt{r}_G$ to ensure

$$I\!P\Big(\big\| W\big(\boldsymbol{\vartheta}_G - \breve{\boldsymbol{\theta}}_G\big) \big\| \geq \mathtt{r}_G \,\big|\, \boldsymbol{Y} \Big) = I\!P\big(\|\boldsymbol{\xi}\| > \mathtt{r}_G \big) = \alpha$$

with $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma_G)$, and $\Sigma_G = W D_G^{-2} W^{\top}$.

Can one can use the credible set $\mathcal{E}_G(\mathtt{r})$ as a frequentist confidence set for the true parameter $\boldsymbol{\theta}^*$? The frequentist coverage probability of $\boldsymbol{\theta}^*$ is given by

$$I\!P\big(\boldsymbol{\theta}^* \in \mathcal{E}_G(\mathtt{r})\big) = I\!P\big(\|W(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G)\| \leq \mathtt{r}_G \big).$$

The aim is to show that the the latter is close to $1 - \alpha$.

Existing results cover only very special cases; see e.g.

- [Johnstone, 2010],
- [Bontemps, 2011],
- [Panov and Spokoiny, 2015],
- [Castillo, 2012],
- [Castillo and Nickl, 2013]
- [Belitser, 2017]

and references therein.

Most of the mentioned results are about asymptotic consistency without rate.

For the posterior mean $\breve{\boldsymbol{\theta}}_G = \bigl(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2\bigr)^{-1}\boldsymbol{\Psi}\boldsymbol{Y}$, it holds

$$I\!\!E\bigl[W\bigl(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G\bigr)\bigr] = W\bigl(I - \Pi_G\bigr)\boldsymbol{\theta}^* \stackrel{\text{def}}{=} \boldsymbol{a}$$

with $\Pi_G = \bigl(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + \sigma^2 G^2\bigr)^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top$. Further,

$$\begin{aligned}\Sigma &\stackrel{\text{def}}{=} \operatorname{Var}\bigl\{W\bigl(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G\bigr)\bigr\} \\ &= W\bigl(\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2\bigr)^{-1}\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top\bigl(\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2\bigr)^{-1}W^\top.\end{aligned}$$

Hence, the vector $W\bigl(\boldsymbol{\theta}^* - \breve{\boldsymbol{\theta}}_G\bigr)$ is under $I\!\!P$ normal with mean $\boldsymbol{a} = W\bigl(I - \Pi_G\bigr)\boldsymbol{\theta}^*$ and variance Σ. Therefore, for $\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma)$

$$I\!\!P\bigl(\boldsymbol{\theta}^* \in \mathcal{E}_G(\mathtt{r})\bigr) = I\!\!P\bigl(\bigl\|\boldsymbol{a} + \boldsymbol{\xi}\bigr\| \le \mathtt{r}\bigr).$$

So, it remains to compare two probabilities

$$I\!\!P\bigl(\|\boldsymbol{a}+\boldsymbol{\xi}\| \le \mathtt{r}\bigr) \text{ vs } I\!\!P\bigl(\|\boldsymbol{\xi}_G\| \le \mathtt{r}\bigr), \qquad \mathtt{r} > 0.$$

Obviously $\Sigma \le \Sigma_G$ for

$$\Sigma = W\bigl(\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2\bigr)^{-1}\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top\bigl(\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2\bigr)^{-1}W^\top$$

$$\Sigma_G = W\bigl(\sigma^{-2}\boldsymbol{\Psi}\boldsymbol{\Psi}^\top + G^2\bigr)^{-1}W^\top$$

yielding

$$\bigl\|\Sigma_G - \Sigma\bigr\|_1 = \operatorname{tr}\Sigma_G - \operatorname{tr}\Sigma\,.$$

Our bound:

$$\bigl|I\!\!P\bigl(\boldsymbol{\theta}^* \notin \mathcal{E}_G(\mathtt{r}_G)\bigr) - \alpha\bigr| \le \frac{\mathtt{C}\bigl(\operatorname{tr}\Sigma_G - \operatorname{tr}\Sigma + \|\boldsymbol{a}\|^2\bigr)}{\|\Sigma_G\|_{\mathrm{Fr}}}\,.$$

We denote by $\Lambda_{m,\boldsymbol{\xi}}$ the partial Frobenius norm of $\Sigma_{\boldsymbol{\xi}}$ defined by

$$\Lambda_{m,\boldsymbol{\xi}}^2 \stackrel{\text{def}}{=} \sum_{k=m}^{\infty} \lambda_k^2(\Sigma_{\boldsymbol{\xi}}), \qquad m = 1, 2.$$

Theorem

Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be Gaussian elements in $\mathcal{H}$ with zero mean and covariance operators $\Sigma_{\boldsymbol{\xi}}$ and $\Sigma_{\boldsymbol{\eta}}$ respectively. For any $\boldsymbol{a} \in \mathcal{H}$

$$\sup_{x>0} \left| I\!\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!\!P(\|\boldsymbol{\eta} - \boldsymbol{a}\| \le x) \right| \lesssim \mathtt{C}_0 \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1,$$

$$\mathtt{C}_0 \stackrel{\text{def}}{=} \frac{1}{\sqrt{\Lambda_{1,\boldsymbol{\xi}} \Lambda_{2,\boldsymbol{\xi}}}} \left(1 + \frac{\|\boldsymbol{a}\|^2}{\Lambda_{2,\boldsymbol{\xi}}}\right) + \frac{1}{\sqrt{\Lambda_{1,\boldsymbol{\eta}} \Lambda_{2,\boldsymbol{\eta}}}} \left(1 + \frac{\|\boldsymbol{a}\|^2}{\Lambda_{2,\boldsymbol{\eta}}}\right).$$

Theorem

Let $\boldsymbol{\xi}$ *be the Gaussian element in* $\mathcal{H}$ *with zero mean and covariance operator* Σ. *For any* $\boldsymbol{a} \in \mathcal{H}$

$$\sup_{x>0} \left| I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\xi}\| \le x) \right| \lesssim \frac{\|\boldsymbol{a}\|^2}{\sqrt{\Lambda_{1,\boldsymbol{\xi}} \Lambda_{2,\boldsymbol{\xi}}}}$$

with

$$\Lambda_{m,\boldsymbol{\xi}}^2 \stackrel{\text{def}}{=} \sum_{k=m}^{\infty} \lambda_k^2(\Sigma_{\boldsymbol{\xi}}), \qquad m = 1, 2.$$

The Pinsker bound yields

$$\sup_{x>0} \left| I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - I\!P(\|\boldsymbol{\xi}\| \le x) \right| \lesssim \|\Sigma^{-1/2} \boldsymbol{a}\|.$$

Corollary

Under conditions of Theorem 1 we have

$$\sup_{x>0} \left| \mathbb{P}(\|\boldsymbol{\xi} - \boldsymbol{a}\| \le x) - \mathbb{P}(\|\boldsymbol{\eta}\| \le x) \right| \lesssim \mathtt{C}_1 \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 + \frac{\|\boldsymbol{a}\|^2}{\sqrt{\Lambda_{1,\boldsymbol{\xi}} \Lambda_{2,\boldsymbol{\xi}}}},$$

$$\textit{where} \quad \mathtt{C}_1 \stackrel{\text{def}}{=} \frac{1}{\sqrt{\Lambda_{1,\boldsymbol{\xi}} \Lambda_{2,\boldsymbol{\xi}}}} + \frac{1}{\sqrt{\Lambda_{1,\boldsymbol{\eta}} \Lambda_{2,\boldsymbol{\eta}}}} .$$

The Pinsker bound yields

$$\sup_{x>0} \left| \mathbb{P}(\|\boldsymbol{\eta} - \boldsymbol{a}\| \le x) - \mathbb{P}(\|\boldsymbol{\xi}\| \le x) \right|$$

$$\lesssim \|\Sigma_{\boldsymbol{\xi}}^{-1/2} \Sigma_{\boldsymbol{\eta}} \Sigma_{\boldsymbol{\xi}}^{-1/2} - I_p\|_{\mathrm{Fr}} + \|\Sigma_{\boldsymbol{\xi}}^{-1/2} \boldsymbol{a}\|.$$

Theorem

Let ξ *be a Gaussian element in* $\mathcal{H}$ *with zero mean and covariance operator* Σ. *Then for arbitrary* $\varDelta > 0$ *and any* $\lambda > \lambda_1$

$$\mathbb{P}(x < \|\boldsymbol{\xi}\|^2 < x + \varDelta) \lesssim \mathtt{C}_2\, \varDelta,$$

where

$$\mathtt{C}_2 \stackrel{\text{def}}{=} \sup_x p_{\boldsymbol{\xi}}(x) \le \frac{1}{\sqrt{\varLambda_{1,\boldsymbol{\xi}}\, \varLambda_{2,\boldsymbol{\xi}}}}$$

and $p_{\boldsymbol{\xi}}$ *stands for the density of* $\|\boldsymbol{\xi}\|^2$. *In particular,*

$$\sup_{x>0} \mathbb{P}(x < \|\boldsymbol{\xi}\|^2 < x + \varDelta) \lesssim \frac{\varDelta}{\sqrt{\varLambda_{1,\boldsymbol{\xi}}\, \varLambda_{2,\boldsymbol{\xi}}}}.$$

Outline

Fix any $s : 0 \le s \le 1$. Let $Z(s)$ be a Gaussian random element in $\mathcal{H}$ with zero mean and covariance operator $\boldsymbol{V}(s)$:

$$\boldsymbol{V}(s) \stackrel{\text{def}}{=} s\Sigma_{\boldsymbol{\xi}} + (1-s)\Sigma_{\boldsymbol{\eta}}.$$

Denote by $\lambda_1(s) \ge \lambda_2(s) \ge \ldots$ the eigenvalues of $\boldsymbol{V}(s)$ and introduce the operator $\boldsymbol{G}(t,s) \stackrel{\text{def}}{=} (\boldsymbol{I} - 2it\,\boldsymbol{V}(s))^{-1}$. Then

$$\begin{aligned} f(t,s,\boldsymbol{a}) &= \mathbb{E}\exp\{it\|Z(s)-\boldsymbol{a}\|^2\} \\ &= \exp\left\{it\left(\|\boldsymbol{a}\|^2 + \langle \boldsymbol{G}(t,s)\boldsymbol{a},\boldsymbol{a}\rangle - \frac{1}{2it}\operatorname{tr}\log\bigl(\boldsymbol{I} - 2it\boldsymbol{V}(s)\bigr)\right)\right\}, \end{aligned}$$

where for an operator $\boldsymbol{A}$ and the identity operator I we use notation

$$\log(\boldsymbol{I} + \boldsymbol{A}) = \boldsymbol{A}\int_0^1 (\boldsymbol{I} + y\boldsymbol{A})^{-1}dy.$$

For a continues d.f. $F(x)$ with c.f. $f(t)$

$$F(x) = \frac{1}{2} + \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_{|t|\le T} e^{-itx} f(t) \frac{dt}{t}.$$

Let us fix an arbitrary $x > 0$. Then

$$\begin{aligned} &\mathbb{P}(\|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x) - \mathbb{P}(\|\boldsymbol{\eta} - \boldsymbol{a}\|^2 < x) \\ &= \frac{i}{2\pi} \lim_{T\to\infty} \text{V.P.} \int_{|t|\le T} \frac{f(t,1,\boldsymbol{a}) - f(t,0,\boldsymbol{a})}{t} e^{-itx}\, dt. \end{aligned}$$

By the Newton-Leibnitz formula

$$f(t,1,\boldsymbol{a}) - f(t,0,\boldsymbol{a}) = \int_0^1 \frac{\partial f(t,s,\boldsymbol{a})}{\partial s}\, ds.$$

Further

$$\frac{\partial f(t,s,\boldsymbol{a})/\partial s}{t} = f(t,s,\boldsymbol{a}) \times \Big[i \, \mathrm{tr}\big\{ (\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}) \boldsymbol{G}(t,s) \big\} - t \langle \boldsymbol{G}(t,s)(\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}) \boldsymbol{G}(t,s) \boldsymbol{a}, \boldsymbol{a} \rangle \Big].$$

Changing the order of integration yields

$$\begin{aligned} & I\!P(\|\boldsymbol{\xi} - \boldsymbol{a}\|^2 < x) - I\!P(\|\boldsymbol{\eta} - \boldsymbol{a}\|^2 < x) \\ & = -\frac{1}{2\pi} \int_0^1 \int_{-\infty}^{\infty} \Big[\mathrm{tr} \left\{ (\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}) \boldsymbol{G}(t,s) \right\} \\ & \qquad + it \langle \boldsymbol{G}(t,s)(\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}) \boldsymbol{G}(t,s) \boldsymbol{a}, \boldsymbol{a} \rangle \Big] f(t,s,\boldsymbol{a}) e^{-itx} \, dt \, ds. \end{aligned}$$

Let us fix s and consider the following quantity

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\mu_j(t,s)f(t,s,\boldsymbol{a})e^{-itx}\,dt,$$

where $\mu_j(t,s) \stackrel{\text{def}}{=} (1-2it\lambda_j(s))^{-1}$ are the eigenvalues of $\boldsymbol{G}(t,s)$. Let also $\overline{Z}_j(s)$ be a random variable with exponential distribution $\mathrm{Exp}(0,1/(2\lambda_j(s)))$, which is independent of $Z_k, k\geq 1$. Then

$$\mathbb{E}e^{it\overline{Z}_j(s)} = \mu_j(t,s).$$

Moreover, $\mu_j(t,s)f(t,s,\boldsymbol{a})$ is the characteristic function of $\overline{Z}_j(s)+\|Z(s)-a\|^2$. Let $p_j(x,s,a)$ be a density function corresponding to $\mu_j(t,s)f(t,s,\boldsymbol{a})$. Then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\mu_j(t,s)f(t,s,\boldsymbol{a})e^{-itx}\,dt = p_j(x,s,a).$$

Denote by $\boldsymbol{P}(x,s,a)$ a diagonal operator with $p_j(x,s,a)$ on the main diagonal. It follows

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{tr}\left\{(\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}})\boldsymbol{G}(t,s)\right\} f(t,s,\boldsymbol{a})e^{-itx}\,dt$$
$$= \mathrm{tr}\left\{(\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}})\boldsymbol{U}^{\top}\boldsymbol{P}(x,s,a)\boldsymbol{U}\right\} \le \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 \|\boldsymbol{P}(x,s,a)\|$$
$$= \|\Sigma_{\boldsymbol{\xi}} - \Sigma_{\boldsymbol{\eta}}\|_1 \; \max_j p_j(x,s,a).$$

$p_j(x,s,a)$ is the density for the c.f. $\mu_j(t,s)f(t,s,\boldsymbol{a})$. Hence,

$$p_j(x,s,a) \le g(x,s,\boldsymbol{a})$$

where $g(x,s,\boldsymbol{a})$ is the density of $\|\boldsymbol{\xi}(s) - \boldsymbol{a}\|^2$.

Require an approximation for the probability of a ball in the form

$$\sup_{x \geq 0} \left| I\!P\left(\|\boldsymbol{S}\|^2 \leq x \right) - I\!P\left(\|\widetilde{\boldsymbol{S}}\|^2 \leq x \right) \right|,$$

where $\widetilde{\boldsymbol{S}} \sim \mathcal{N}(0, \Sigma)$.

[Bentkus, 2003]: in the i.i.d. case

$$\sup_{A \in \mathcal{C}} \left| I\!P\left(\boldsymbol{S} \in A \right) - I\!P\left(\widetilde{\boldsymbol{S}} \in A \right) \right| \lesssim \sqrt{p^3/n}$$

Here $\mathcal{C}$ is the class of all convex sets in $I\!R^p$.

Our guess:

$$\sup_{x \geq 0} \left| I\!P\left(\|\boldsymbol{S}\|^2 \leq x \right) - I\!P\left(\|\widetilde{\boldsymbol{S}}\|^2 \leq x \right) \right| \lesssim \frac{\operatorname{tr}^{3/2}(\Sigma)}{n} .$$

[Koltchinskii and Lounici, 2015] studied a quality of estimation of the spectral projector of the sample covariance.

[Naumov et al., 2017]: a bootstrap procedure for building a confidence set for the target projector;

[Silin and Sp, 2017]: nonparametric Bayes approach for the spectral projector

Belitser, E. (2017).

On coverage and local radial rates of credible sets.

Ann. Statist., 45(3):1124–1151.

Bentkus, V. (2003).

A new approach to approximations in probability theory and operator theory.

Liet. Mat. Rink., 43(4):444–470.

Bontemps, D. (2011).

Bernstein-von Mises theorems for Gaussian regression with increasing number of regressors.

Ann. Statist., 39(5):2557–2584.

Castillo, I. (2012).

A semiparametric Bernstein - von Mises theorem for Gaussian process priors.

Probability Theory and Related Fields, 152:53–99.

10.1007/s00440-010-0316-5.

Castillo, I. and Nickl, R. (2013).

Nonparametric Bernstein–von Mises theorems in Gaussian white noise.

Ann. Statist., 41(4):1999–2028.

Johnstone, I. M. (2010).

High dimensional Bernstein–von Mises: simple examples.

In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*, volume 6 of *Inst. Math. Stat. Collect.*, pages 87–98. Inst. Math. Statist., Beachwood, OH.

Koltchinskii, V. and Lounici, K. (2015).

Normal approximation and concentration of spectral projectors of sample covariance.

arXiv:1504.07333.

Naumov, A., Spokoiny, V., and Ulyanov, V. (2017).

Bootstrap confidence sets for spectral projectors of sample covariances.

arXiv:1703.00871.

Panov, M. and Spokoiny, V. (2015).

Finite Sample Bernstein - von Mises Theorem for Semiparametric Problems.

Bayesian Analysis, 10(3):665–710.

Spokoiny, V. and Zhilova, M. (2015).
Bootstrap confidence sets under model misspecification.
Ann. Statist., 43(6):2653–2675.
arXiv:1410.0347.