Exercises to VL WS2014

Vladimir Spokoiny*

Weierstrass-Institute,
Humboldt University Berlin,
Moscow Institute of Physics and Technology
Mohrenstr. 39, 10117 Berlin, Germany,

spokoiny@wias-berlin.de

January 30, 2015

Abstract

AMS 2000 Subject Classification: Primary . Secondary
Keywords:

*

21.10

Exercise 0.1. Consider the regression model

$$Y_i = f(X_i, \boldsymbol{\theta}^*) + \varepsilon_i \qquad \mathbb{E}\varepsilon_i = 0.$$

with independent zero-mean errors ε_i: $\mathbb{E}\varepsilon_i = 0$.

Prove that for any $\boldsymbol{\theta}$

$$\mathbb{E}_{\boldsymbol{\theta}^*} \sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}) \bigr|^2 \ge \mathbb{E}_{\boldsymbol{\theta}^*} \sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}^*) \bigr|^2,$$

that is, $\boldsymbol{\theta}^*$ minimizes over $\boldsymbol{\theta}$ the expectation under the true measure of the sum $\sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}) \bigr|^2$.

Hint: Observe that $\varepsilon_i = Y_i - f(X_i, \boldsymbol{\theta}^*)$ and hence, the latter r.v. has mean zero. We now use the following simple fact: if $\mathbb{E}(\varepsilon) = 0$, then for any $z \ne 0$

$$\mathbb{E}(\varepsilon + z)^2 \ge \mathbb{E}\varepsilon^2.$$

Exercise 0.2. Consider the median regression model

$$Y_i = f(X_i, \boldsymbol{\theta}^*) + \varepsilon_i, \qquad \operatorname{med}(\varepsilon_i) = 0.$$

Prove that for any $\boldsymbol{\theta}$

$$\mathbb{E}_{\boldsymbol{\theta}^*} \sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}) \bigr| \ge \mathbb{E}_{\boldsymbol{\theta}^*} \sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}^*) \bigr|,$$

that is, $\boldsymbol{\theta}^*$ minimizes over $\boldsymbol{\theta}$ the expectation under the true measure of the sum $\sum \bigl| Y_i - f(X_i, \boldsymbol{\theta}) \bigr|$.

Hint: Observe that $\varepsilon_i = Y_i - f(X_i, \boldsymbol{\theta}^*)$ and hence, the latter r.v. has median zero. Use the following simple fact: if $\operatorname{med}(\varepsilon) = 0$, then for any $z \ne 0$

$$\mathbb{E}|\varepsilon + z| \ge \mathbb{E}|\varepsilon|.$$

Exercise 0.3. Extend the result of Exercise 0.6 to quantile regression. If $\mathbb{P}(\varepsilon_i \ge 0) = \tau$, then for any $\boldsymbol{\theta}$

$$\mathbb{E}_{\boldsymbol{\theta}^*} \sum g_\tau\bigl(Y_i - f(X_i, \boldsymbol{\theta}) \bigr) \ge \mathbb{E}_{\boldsymbol{\theta}^*} \sum g_\tau\bigl(Y_i - f(X_i, \boldsymbol{\theta}^*) \bigr),$$

with $g_\tau(x) = -\tau x \, \mathbb{I}(x \le 0) + (1 - \tau) x \, \mathbb{I}(x > 0)$.

Exercise 0.4. Consider the nonparametric regression model

$$Y_i = f(X_i) + \varepsilon_i, \qquad \mathbb{E}\varepsilon_i = 0$$

Show that the nonparametric LS-estimator $\widetilde{f}$ with

$$\widetilde{f} = \operatorname*{argmin}_{f \in \mathcal{F}_n} \sum_i |Y_i - f(X_i)|^2$$

where $\mathcal{F}_n$ is the class of all functions given in the design points X_i, coincides with the observations $\boldsymbol{Y}$:

$$\widetilde{f}(X_i) = Y_i .$$

28.10

Exercise 0.5. Let the regressor $\boldsymbol{x}$ be d-dimensional, $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$. Describe the basis system and the corresponding vector of coefficients for the case when $f(\boldsymbol{x})$ is a quadratic function of x.

Exercise 0.6. Let $\{\psi_m\}$ be an orthonormal polynomial system. Show that for any polynomial $P_j(x)$ of degree $j < m$, it holds $\langle P_j, \psi_m \rangle = 0$.

The Fourier basis is composed by the constant function $F_0 \equiv 1$ and the functions $F_{2m-1}(x) = \sin(2m\pi x)$ and $F_{2m}(x) = \cos(2m\pi x)$ for $m = 1, 2, \ldots$.

The cosine basis is composed by the functions $S_0 \equiv 1$, and $S_m(x) = \cos(m\pi x)$ for $m \geq 1$.

Trigonometric identities imply orthogonality

$$\bigl\langle F_m, F_j \bigr\rangle = \int_0^1 F_m(x) F_j(x) dx = 0, \qquad j \neq m. \tag{0.1}$$

Also

$$\int_0^1 F_m^2(x) dx = 1/2 \tag{0.2}$$

Exercise 0.7. Check (0.1) and (0.2). Also check that

$$\int_0^1 S_j(x) S_m(x) dx = \frac{1}{2}\, \mathbb{1}(j = m).$$

Exercise 0.8. With $u = \cos(\pi x)$, it holds $S_m(x) = T_m(u)$, where $T_m(x)$ is the mth Chebyshev polynomial. Show that any expansion of the function $f(u)$ by the Chebyshev polynomials yields an expansion of $f\bigl(\cos(\pi x)\bigr)$ by the cosine system.

11.04

Exercise 0.9. Consider a non-overlapping partition of $\mathcal{X}$ into intervals A_k for $k = 1, \ldots, K$.

Check that for each $k \le K$, there exists a point x_k such that

$$\sum (X_i - x_k)\, \mathbb{I}(X_i \in A_k) = 0. \tag{0.3}$$

Introduce for each $k \le K$ two basis functions $\phi_{j-1}(x) = \mathbb{I}(x \in A_k)$ and $\phi_j(x) = (x - x_k)\, \mathbb{I}(x \in A_k)$ with $j = 2k$.

Assume (0.3) for each $k \le K$. Check that any piecewise linear function can be uniquely represented in the form

$$f(x, \boldsymbol{\theta}) = \sum_{j=1}^{p} \theta_j \phi_j(x)$$

with $p = 2K$ and the functions ϕ_j are orthogonal in the sense that for $j \ne j'$

$$\sum_{i=1}^{n} \phi_j(X_i)\phi_{j'}(X_i) = 0.$$

In addition, for each $k \le K$

$$\|\phi_j\|^2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \phi_j^2(X_i) = \begin{cases} N_k, & j = 2k-1, \\ V_k^2 & j = 2k. \end{cases}$$

$$N_k^2 \stackrel{\text{def}}{=} \sum_{X_i \in A_k} 1, \qquad V_k^2 \stackrel{\text{def}}{=} \sum_{X_i \in A_k} (X_i - x_k)^2.$$

Check that in the case of a homogeneous Gaussian regression $Y_i = f(X_i, \boldsymbol{\theta}) + \varepsilon_i$ with ε_i iid $\mathcal{N}(0, \sigma^2)$, orthogonality of the basis helps to gain a simple closed form for the MLE estimator $\widetilde{\boldsymbol{\theta}} = (\widetilde{\theta}_j)$:

$$\widetilde{\theta}_j = \frac{1}{\|\psi_j\|^2} \sum_{i=1}^{n} Y_i \psi_j(X_i) = \begin{cases} \frac{1}{N_k} \sum_{X_i \in A_k} Y_i\,, & j = 2k-1, \\ \frac{1}{V_k^2} \sum_{X_i \in A_k} Y_i (X_i - x_k), & j = 2k. \end{cases}$$

Exercise 0.10. Let $f(x)$ be a q-spline with the support on $q' < q$ neighbor spans A_k, A_{k+1}, $\ldots$ $A_{k+q'-1}$. Then $f(x) \equiv 0$.
Hint: consider any spline of the form $f(x) = \sum_{j=k}^{k+q'-1} c_j \phi_j(x)$. Show that the boundary conditions $f^{(m)}(t_{k+q'}) = 0$ for $m = 0, 1, \ldots, q$ yield $c_j \equiv 0$.

Exercise 0.11. The basis B–spline functions can be constructed successfully. For $q = 0$, $b_{k,0}(x) = \mathbb{I}(x \in A_k)$, $k = 1, \ldots, K$. Each linear B–spline $b_{k,1}(x)$ has a triangle shape on the two connected intervals A_k and A_{k+1}. It can be defined by the formula

$$b_{k,1}(x) \stackrel{\text{def}}{=} \frac{x - t_{k-1}}{t_k - t_{k-1}} b_{k,0}(x) + \frac{t_{k+1} - x}{t_{k+1} - t_k} b_{k+1,0}(x), \quad k = 1, \ldots, K - 1.$$

One can continue this way leading to the *Cox–de Boor recursion formula*

$$b_{k,m}(x) \stackrel{\text{def}}{=} \frac{x - t_{k-1}}{t_{k+m-1} - t_{k-1}} b_{k,m-1}(x) + \frac{t_{k+m} - x}{t_{k+m} - t_k} b_{k+1,m-1}(x)$$

for $k = 1, \ldots, K - m$.

Check by induction for each function $b_{k,m}(x)$ the following conditions:

1. $b_{k,m}(x)$ a polynomial of degree m on each span $A_k, \ldots, A_{k+m-1}$ and zero outside;
2. $b_{k,m}(x)$ can be uniquely represented as a sum $b_{k,m}(x) = \sum_{l=0}^{m-1} c_{l,k} \phi_{k+l}(x)$;
3. $b_{k,m}(x)$ is a m-spline.

Exercise 0.12. Let Y_i be binary outputs. Consider the generalized regression model

$$Y_i \sim P_{f(X_i)} \in (P_{\boldsymbol{\upsilon}})$$

where $(P_{\boldsymbol{\upsilon}}, \boldsymbol{\upsilon} \in [0, 1])$ is a Bernoulli family: $P_{\boldsymbol{\upsilon}}(Y = 1) = \boldsymbol{\upsilon}$. Describe the model and the MLE $\widetilde{\boldsymbol{\theta}}$ for a piecewise constant regression function f.

Exercise 0.13. (*) Consider the linear logit regression

$$L(\boldsymbol{\theta}) = \sum_i \bigl[Y_i \Psi_i^\top \boldsymbol{\theta} - \log(1 + \mathrm{e}^{\Psi_i^\top \boldsymbol{\theta}})\bigr].$$

The corresponding estimate reads as

$$\widetilde{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) = \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_i \bigl[Y_i \Psi_i^\top \boldsymbol{\theta} - \log(1 + \mathrm{e}^{\Psi_i^\top \boldsymbol{\theta}})\bigr]$$

- Specify the estimating equation for the case of logit regression.
- Specify the step of the Newton-Raphson procedure for computing the MLE $\widetilde{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$.

11.11

Exercise 0.14. Consider univariate polynomial regression of degree $p-1$. This means that f is a polynomial function of degree $p-1$ observed at the points X_i with errors ε_i that are assumed to be i.i.d. normal. The function f can be represented as

$$f(x) = \theta_1^* + \theta_2^* x + \ldots + \theta_p^* x^{p-1}$$

using the basis functions $\psi_m(x) = x^{m-1}$ for $m = 0, \ldots, p-1$. At the same time, for any point x_0, this function can also be written as

$$f(x) = u_1^* + u_2^*(x - x_0) + \ldots + u_p^*(x - x_0)^{p-1}$$

using the basis functions $\breve{\psi}_m = (x - x_0)^{m-1}$.

- Write the matrices Ψ and $\Psi\Psi^\top$ and similarly $\breve{\Psi}$ and $\breve{\Psi}\breve{\Psi}^\top$.
- Describe the linear transformation A such that $\boldsymbol{u} = A\boldsymbol{\theta}$ for $p = 1$.
- Describe the transformation A such that $\boldsymbol{u} = A\boldsymbol{\theta}$ for $p > 1$.

Hint: use the formula

$$u_m^* = \frac{1}{(m-1)!} f^{(m-1)}(x_0), \qquad m = 1, \ldots, p$$

to identify the coefficient u_m^* via $\theta_m^*, \ldots, \theta_p^*$.

Exercise 0.15. Consider the case of an orthogonal design with $\Psi\Psi^\top = \boldsymbol{I}_p$. Specify the projector $\Pi = \Psi^\top \bigl(\Psi\Psi^\top\bigr)^{-1}\Psi$ for this situation, particularly its decomposition

$$\Pi = U^\top \Lambda_p U$$

where U is an orthonormal matrix and Λ_p is a diagonal matrix with the first p diagonal elements equal to 1 and the others equal to zero:

$$\Lambda_p = \operatorname{diag}\{\underbrace{1, \ldots, 1}_{p}, \underbrace{0, \ldots, 0}_{n-p}\}.$$

Exercise 0.16. Consider the linear model $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$ with $\operatorname{Var}(\boldsymbol{\varepsilon}) = \Sigma$. Compute

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

Exercise 0.17. Check that the linear transformation $\check{\boldsymbol{Y}} = \Sigma^{-1/2}\boldsymbol{Y}$ of the data does not change the value of the log-likelihood ratio $L(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = L(\boldsymbol{\theta}) - L(\boldsymbol{\theta}^*)$ and hence, of the maximum likelihood $L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*)$.
Hint: use the representation

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{1}{2}(\boldsymbol{Y} - \Psi^\top\boldsymbol{\theta})^\top\Sigma^{-1}(\boldsymbol{Y} - \Psi^\top\boldsymbol{\theta}) + R \\ &= \frac{1}{2}(\check{\boldsymbol{Y}} - \check{\Psi}^\top\boldsymbol{\theta})^\top(\check{\boldsymbol{Y}} - \check{\Psi}^\top\boldsymbol{\theta}) + R \end{aligned}$$

and check that the transformed data $\check{\boldsymbol{Y}}$ is described by the model $\check{\boldsymbol{Y}} = \check{\Psi}^\top\boldsymbol{\theta}^* + \check{\boldsymbol{\varepsilon}}$ with $\check{\Psi} = \Psi\Sigma^{-1/2}$ and $\check{\boldsymbol{\varepsilon}} = \Sigma^{-1/2}\boldsymbol{\varepsilon} \sim \mathcal{N}(0, I_n)$ yielding the same log-likelihood ratio as in the original model.

Exercise 0.18. Consider the linear model $\boldsymbol{Y} = \Psi^\top\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$ with $\Sigma = \sigma^2 I_n$. Then it holds

$$2L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^*) = \sigma^{-2}\|\Pi\boldsymbol{\varepsilon}\|^2$$

where $\Pi = \Psi^\top\bigl(\Psi\Psi^\top\bigr)^{-1}\Psi$ is the projector in $\mathbb{R}^n$ on the subspace spanned by the vectors $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_p$.
Hint: use that $\boldsymbol{\zeta} = \sigma^{-2}\Psi\boldsymbol{\varepsilon}$, $D^2 = \sigma^{-2}\Psi\Psi^\top$, and

$$\sigma^{-2}\|\Pi\boldsymbol{\varepsilon}\|^2 = \sigma^{-2}\boldsymbol{\varepsilon}^\top\Pi^\top\Pi\boldsymbol{\varepsilon} = \sigma^{-2}\boldsymbol{\varepsilon}^\top\Pi\boldsymbol{\varepsilon} = \boldsymbol{\zeta}^\top D^{-2}\boldsymbol{\zeta}.$$

11.18

Exercise 0.19. Consider a linear regression model

$$Y_i = \sum_{j=1}^{p} \theta_j \psi_j(X_i) + \varepsilon_i \tag{0.4}$$

for $i = 1, \ldots, n$ with the design-orthonormal functions ψ_j :

$$\sum_{i=1}^{n} \psi_j(X_i)\, \psi_{j'}(X_i) = \delta_{jj'} .$$

Show for the LSE $\widetilde{\boldsymbol{\theta}}^{*}$ and the corresponding function estimate $\widetilde{f}(X_i) = \sum_{j=1}^{p} \widetilde{\theta}_j \psi_j(X_i)$ that

$$I\!E \|\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^{*}\|^2 = I\!E \|\widetilde{f} - f\|^2 = I\!E \sum_{i} \bigl| \widetilde{f}(X_i) - f(X_i) \bigr|^2$$

Exercise 0.20. Let $L(\boldsymbol{\theta})$ be the log-likelihood for the linear regression model (0.4). Compute the negative Hessian matrix $D^2 = -\nabla^2 I\!E L(\boldsymbol{\theta})$. Show that it is deterministic and does not depend on the data distribution. Check what happens with this matrix if the model assumptions are misspecified.

Exercise 0.21. Let $L(\boldsymbol{\theta})$ be the log-likelihood for the linear regression model (0.4). Check that the linear transformation $\check{\boldsymbol{Y}} = \Sigma^{-1/2} \boldsymbol{Y}$ of the data does not change the value of the log-likelihood ratio $L(\boldsymbol{\theta}, \boldsymbol{\theta}^{*})$ and hence, of the maximum likelihood $L(\widetilde{\boldsymbol{\theta}}, \boldsymbol{\theta}^{*})$. Hint: use the representation

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{1}{2} (\boldsymbol{Y} - \Psi^{\top} \boldsymbol{\theta})^{\top} \Sigma^{-1} (\boldsymbol{Y} - \Psi^{\top} \boldsymbol{\theta}) + R \\ &= \frac{1}{2} (\check{\boldsymbol{Y}} - \check{\Psi}^{\top} \boldsymbol{\theta})^{\top} (\check{\boldsymbol{Y}} - \check{\Psi}^{\top} \boldsymbol{\theta}) + R \end{aligned}$$

and check that the transformed data $\check{\boldsymbol{Y}}$ is described by the model $\check{\boldsymbol{Y}} = \check{\Psi}^{\top} \boldsymbol{\theta}^{*} + \check{\boldsymbol{\varepsilon}}$ with $\check{\Psi} = \Psi \Sigma^{-1/2}$ and $\check{\boldsymbol{\varepsilon}} = \Sigma^{-1/2} \boldsymbol{\varepsilon} \sim \mathcal{N}(0, I_n)$ yielding the same log-likelihood ratio as in the original model.

Exercise 0.22. Let $D^2 = \Psi \Sigma^{-1} \Psi^{\top}$. Check that the likelihood-based CS $\mathcal{E}(\mathfrak{z}_\alpha)$ and estimate-based CS $E(z_\alpha) = \{\boldsymbol{\theta} : \|D(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})\| \le z_\alpha\}$, $z_\alpha^2 = 2\mathfrak{z}_\alpha$, coincide in the case of the linear modeling:

$$\mathcal{E}(\mathfrak{z}_\alpha) = \bigl\{ \boldsymbol{\theta} : \bigl\| D(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \bigr\|^2 \le 2\mathfrak{z}_\alpha \bigr\}.$$

Exercise 0.23. Show that an overestimation of the noise in the sense $\Sigma \ge \Sigma_0$ preserves the coverage probability for the CS $\mathcal{E}(\mathfrak{z}_\alpha)$, that is, if $2\mathfrak{z}_\alpha$ is the $1 - \alpha$ quantile of χ_p^2, then $I\!P\bigl(\mathcal{E}(\mathfrak{z}_\alpha) \not\ni \boldsymbol{\theta}^{*}\bigr) \le \alpha$.

11.25

Exercise 0.24. Consider the regression model

$$Y_i = \theta_1^* \psi_1(X_i) + \ldots \theta_p^* \psi_p(X_i) + \varepsilon_i \tag{0.5}$$

with independent heterogeneous errors $\operatorname{Var}(\varepsilon_i) = \sigma_i^2$. Consider the MLE $\widetilde{\boldsymbol{\theta}}$ and the LSE $\widetilde{\boldsymbol{\theta}}_{LSE} = \bigl(\Psi\Psi^\top\bigr)^{-1}\Psi \boldsymbol{Y}$ and corresponding to homogeneous errors.

1. Compute $\widetilde{\boldsymbol{\theta}}$
2. Show that $\mathbb{E}\widetilde{\boldsymbol{\theta}} = \mathbb{E}\widetilde{\boldsymbol{\theta}}_{LSE} = \boldsymbol{\theta}^*$.
3. Compute the variance $\operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}\bigr)$ and the variance $\operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}_{LSE}\bigr)$;
4. show that $\operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}_{LSE}\bigr) \ge \operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}\bigr)$;
5. check that $\operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}_{LSE}\bigr) = \operatorname{Var}\bigl(\widetilde{\boldsymbol{\theta}}\bigr)$ iff all the σ_i are equal to each other.

Exercise 0.25. Consider the nonparametric model

$$Y_i = f(X_i) + \varepsilon_i, \qquad \operatorname{Var}(\varepsilon_i) = \sigma_i^2. \tag{0.6}$$

and the parametric approximation (0.6). Consider the qMLE=LSE $\widetilde{\boldsymbol{f}}_{LSE} = \Pi \boldsymbol{Y}$ for $\Pi = \Psi^\top\bigl(\Psi\Psi^\top\bigr)^{-1}\Psi$.

1. Derive the bias-variance decomposition for the quadratic losses $\|\widetilde{\boldsymbol{f}}_{LSE} - \boldsymbol{f}\|^2$ and of the risk $\mathbb{E}\|\widetilde{\boldsymbol{f}}_{LSE} - \boldsymbol{f}\|^2$.
2. Compute the variance term of $\widetilde{\boldsymbol{f}}_{LSE}$ and of $\widetilde{\boldsymbol{f}} = \Psi^\top \widetilde{\boldsymbol{\theta}}$ for the MLE $\widetilde{\boldsymbol{\theta}}$.

Exercise 0.26. Consider the projection estimator $\widetilde{\boldsymbol{f}}_m = \Pi_m \boldsymbol{Y}$ for the model (0.6) with $\Pi_m = \Psi_m^\top\bigl(\Psi_m\Psi_m^\top\bigr)^{-1}\Psi_m$. For two different values $m' > m$,

1. Check that $\Pi_{m',m} \stackrel{\text{def}}{=} \Pi_{m'} - \Pi_m$ is a projector in $\mathbb{R}^n$. Describe its image in the orthogonal case when $\Psi\Psi^\top$ is a diagonal matrix.
2. Check the identities

$$\|\boldsymbol{Y} - \Pi_m \boldsymbol{Y}\|^2 - \|\boldsymbol{Y} - \Pi_{m'}\boldsymbol{Y}\|^2 = \|\widetilde{\boldsymbol{f}}_{m'} - \widetilde{\boldsymbol{f}}_m\|^2 = \|\Pi_{m',m}\boldsymbol{f} + \Pi_{m',m}\boldsymbol{\varepsilon}\|^2,$$
$$\|\widetilde{\boldsymbol{f}}_{m'} - \boldsymbol{f}\|^2 - \|\widetilde{\boldsymbol{f}}_m - \boldsymbol{f}\|^2 = -\|\Pi_{m',m}\boldsymbol{f}\|^2 + \|\Pi_{m',m}\boldsymbol{\varepsilon}\|^2.$$

3. compute $\mathbb{E}\|\widetilde{\boldsymbol{f}}_{m'} - \widetilde{\boldsymbol{f}}_m\|^2$ and $\mathbb{E}\bigl[\|\widetilde{\boldsymbol{f}}_{m'} - \boldsymbol{f}\|^2 - \|\widetilde{\boldsymbol{f}}_m - \boldsymbol{f}\|^2\bigr]$.

Exercise 0.27. Consider the model (0.6) with homogeneous errors $\sigma_i \equiv \sigma$. Let m^* be the oracle choice

1. check that for $m > m^*$, it holds

$$\|\Pi_{m,m^*} \boldsymbol{f}\|^2 \le \sigma^2 (m - m^*)$$

2. check that for $m < m^*$, it holds

$$\|\Pi_{m^*,m} \boldsymbol{f}\|^2 \ge \sigma^2 (m^* - m).$$

Exercise 0.28. Let

$$\widehat{m} \stackrel{\text{def}}{=} \operatorname*{argmin}_m \|\boldsymbol{Y} - \Pi_m \boldsymbol{Y}\|^2 + 2\sigma^2 m$$

Check the inequality

$$I\!\!P\bigl(\widehat{m} > m^*\bigr) \le \sum_{m > m^*} I\!\!P\bigl(\|\widetilde{\boldsymbol{f}}_m - \widetilde{\boldsymbol{f}}_{m^*}\|^2 > 2(m - m^*)\bigr).$$

02.12

Exercise 0.29. Let $\boldsymbol{\xi}$ be standard Gaussian vector in $\mathbb{R}^k$ and $\boldsymbol{\delta}$ be a deterministic vector in $\mathbb{R}^k$ with $\|\boldsymbol{\delta}\| = \Delta$. Then

- the distribution of $\|\boldsymbol{\xi} + \boldsymbol{\delta}\|^2$ only depends on k and Δ.
- Let, for a given $\mathtt{x}$, the quantiles $\mathfrak{z}^{+}(k, \Delta; \mathtt{x})$ and $\mathfrak{z}^{-}(k, \Delta, \mathtt{x})$ be defined as

$$\mathbb{P}\bigl(\|\boldsymbol{\xi} + \boldsymbol{\delta}\|^2 \ge \mathfrak{z}^{+}(k, \Delta; \mathtt{x})\bigr) = \mathrm{e}^{-\mathtt{x}},$$
$$\mathbb{P}\bigl(\|\boldsymbol{\xi} + \boldsymbol{\delta}\|^2 \le \mathfrak{z}^{-}(k, \Delta; \mathtt{x})\bigr) = \mathrm{e}^{-\mathtt{x}}.$$

Then

$$\mathfrak{z}^{+}(k, \Delta; \mathtt{x}) \le k + \Delta + \mathfrak{z}(k, \mathtt{x}) + 2\Delta^{1/2}\mathfrak{z}_1(\mathtt{x}),$$
$$\mathfrak{z}^{-}(k, \Delta; \mathtt{x}) \ge k + \Delta - \mathfrak{z}(k, \mathtt{x}) - 2\Delta^{1/2}\mathfrak{z}_1(\mathtt{x}),$$

Here $\mathfrak{z}(k, \mathtt{x})$ and $\mathfrak{z}_1(\mathtt{x})$ are quintiles of χ^2_k and of normal r.v. ξ_1:

$$\mathbb{P}(|\xi_1| \ge \mathfrak{z}_1(\mathtt{x})) = \mathrm{e}^{-\mathtt{x}}, \qquad \xi_1 \sim \mathcal{N}(0, 1),$$
$$\mathbb{P}\bigl(\|\boldsymbol{\xi}\|^2 - k > \mathfrak{z}(k, \mathtt{x})\bigr) = \mathrm{e}^{-\mathtt{x}}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, I_k).$$

Exercise 0.30. Let the value of non-centrality parameter Δ be fixed to have $\mathfrak{z}^{\pm}(k, \Delta^{\pm}, \mathtt{x})$ exactly equal to $2k$:

$$\mathfrak{z}^{+}(k, \Delta^{+}(k, \mathtt{x}); \mathtt{x}) = 2k, \qquad \mathfrak{z}^{-}(k, \Delta^{-}(k, \mathtt{x}); \mathtt{x}) = 2k.$$

Equivalently: if $\|\boldsymbol{\delta}^{+}\|^2 \le \Delta^{+}(k, \mathtt{x})$ and $\|\boldsymbol{\delta}^{-}\|^2 \ge \Delta^{-}(k, \mathtt{x})$, then

$$\mathbb{P}\bigl(\|\boldsymbol{\xi} + \boldsymbol{\delta}^{+}\|^2 > 2k\bigr) \le \mathrm{e}^{-\mathtt{x}},$$
$$\mathbb{P}\bigl(\|\boldsymbol{\xi} + \boldsymbol{\delta}^{-}\|^2 < 2k\bigr) \le \mathrm{e}^{-\mathtt{x}}.$$

- Show that $\Delta^{+}(k, \mathtt{x}) < k$, $\Delta^{-}(k, \mathtt{x}) > k$
- Check that

$$\Delta^{+}(k, \mathtt{x}) \ge k - \mathfrak{z}(k, \mathtt{x}) - 2k^{1/2}\mathfrak{z}_1(\mathtt{x}),$$
$$\Delta^{-}(k, \mathtt{x}) \le k + \mathfrak{z}(k, \mathtt{x}) + 2k^{1/2}\mathfrak{z}_1(\mathtt{x}).$$

Exercise 0.31. Let the errors $\boldsymbol{\varepsilon}$ be normal and homogeneous: $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 I_n)$. Then the inequalities

$$\sigma^{-1}\|\boldsymbol{b}_{m,m^*}\| \le \Delta^+(m - m^*, \mathtt{x}), \qquad m > m^*$$
$$\sigma^{-1}\|\boldsymbol{b}_{m^*,m}\| \ge \Delta^-(m^* - m, \mathtt{x}), \qquad m < m^*$$

ensures

$$I\!P\bigl(\widetilde{\mathcal{R}}_m \le \widetilde{\mathcal{R}}_{m^*}\bigr) \le \mathrm{e}^{-\mathtt{x}}.$$

Exercise 0.32. Let $\bigl\{\mathtt{x}_m\bigr\}$ be a set of critical values such that

$$\sum_m \mathrm{e}^{-\mathtt{x}_m} \le \mathrm{e}^{-\mathtt{x}}.$$

If $\mathcal{M}(\mathtt{x})$ is the set of indices m such that

$$\mathcal{M}(\mathtt{x}) \stackrel{\mathrm{def}}{=} \begin{cases} \sigma^{-1}\|\boldsymbol{b}_{m,m^*}\| \ge \Delta^+(m - m^*, \mathtt{x}_m), & m > m^*, \\ \sigma^{-1}\|\boldsymbol{b}_{m^*,m}\| \le \Delta^-(m^* - m, \mathtt{x}_m), & m < m^*, \end{cases}$$

then

$$I\!P\bigl(\widehat{m} \notin \mathcal{M}(\mathtt{x})\bigr) \le \mathrm{e}^{-\mathtt{x}}.$$

Exercise 0.33. Let $\Omega_0(\mathtt{x})$ be a random set such that

$$\|\sigma^{-1}\Pi_{m,m^*}\boldsymbol{\varepsilon}\|^2 - (m - m^*) \le \mathfrak{z}(m - m^*, \mathtt{x}_m), \qquad m > m^*,$$
$$-\|\sigma^{-1}\Pi_{m,m^*}\boldsymbol{\varepsilon}\|^2 + (m - m^*) \le \mathfrak{z}(m^* - m, \mathtt{x}_m), \qquad m < m^*.$$

Show that $I\!P\bigl(\Omega_0(\mathtt{x})\bigr) \ge 1 - \mathrm{e}^{-\mathtt{x}}$ and for every m, the loss $\varrho(\widetilde{\boldsymbol{f}}_m, \boldsymbol{f}^*)$ satisfies on $\Omega_0(\mathtt{x})$

$$\varrho(\widetilde{\boldsymbol{f}}_m, \boldsymbol{f}^*) \ge m - \mathfrak{z}(m, \mathtt{x}).$$

09.12

Exercise 0.34. For the case of an orthonormal design Ψ, find a vector $\boldsymbol{\theta}^*_\alpha$ for which $\sigma^{-2}\|\boldsymbol{b}_{m,m^\circ}\|^2 \equiv \alpha(m-m^\circ)$.

Exercise 0.35. Define

$$\mathtt{z}_\beta(k,\mathtt{x}) \stackrel{\text{def}}{=} \mathfrak{z}^+(k,\alpha k;\mathtt{x}),$$

where $\mathfrak{z}^+(k,\Delta;\mathtt{x})$ is the quantile of a non-central chi-squared.

Check that for each m°,

$$q_{m^\circ}(\alpha) \ge q_{m^\circ+1}(\alpha)$$

yielding

$$\mathtt{z}_\beta(k,\mathtt{x}+q_{m^\circ}) \ge \mathtt{z}_\beta(k,\mathtt{x}+q_{m^\circ-1}).$$

Exercise 0.36. Let $\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\sigma^2 \boldsymbol{I}_n)$, $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$ with $\Psi\Psi^\top = \boldsymbol{I}_p$ and $\boldsymbol{\theta}^* = (\theta^*_1,\ldots,\theta^*_p)$. Let m° and $\mathtt{x}$ be fixed.

- The values $\theta^*_1,\ldots,\theta^*_{m^\circ}$ do not enter in the distribution of T_{m,m°.
- Consider the set $F_\alpha(m^\circ)$ of all vectors $\boldsymbol{f}^*$ with $\sigma^{-2}\big\|\boldsymbol{b}_{m,m^\circ}\big\|^2 \le \alpha(m-m^\circ)$. Within this set, it holds for each $m > m^\circ$

$$I\!\!P\bigl(\bigl\{2T_{m,m^\circ} - \mathtt{z}_\beta(m-m^\circ,\mathtt{x})\bigr\} > 0\bigr) \le \mathrm{e}^{-\mathtt{x}}.$$

Exercise 0.37. Consider the case of a misspecified variance: assumed $\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\sigma^2\boldsymbol{I}_n)$, the truth $\boldsymbol{\varepsilon} \sim \mathcal{N}(0,\sigma_0^2\boldsymbol{I}_n)$. Find $\mathfrak{z}^+(k,\mathtt{x})$ and $\mathfrak{z}^-(k,\mathtt{x})$ which provide for $k = m - m^* > 0$

$$\|\sigma^{-1}\Pi_{m,m^*}\boldsymbol{\varepsilon}\|^2 \le \mathfrak{z}^+(k,\mathtt{x}_m),$$

$$\|\sigma^{-1}\Pi_{m,m^*}\boldsymbol{\varepsilon}\|^2 \le \mathfrak{z}^-(k,\mathtt{x}_m).$$

12.01

Exercise 0.38. The risk of the spectral cut-off estimate $\widetilde{\boldsymbol{u}}_m$ fulfills

$$\mathcal{R}(\widetilde{\boldsymbol{u}}_m) = \sum_{j=1}^{m} \lambda_j^{-1}\sigma^2 + \sum_{j=m+1}^{p} u_j^{*2}.$$

Specify the choice of the oracle cut-off index m^*.

Exercise 0.39. Describe the linear functions for the values

- $f(0.5)$;
- $f''(1)$;
- $\int_0^1 f(x)dx$;
- $\int_0^1 \sin(2\pi x) f'(x)dx$.

Exercise 0.40. Build an example of a signal $\boldsymbol{\theta}^*$ in the sequence space model $Y_i = \theta_j^* + \varepsilon_j$ such that the bias $\|W(\boldsymbol{\theta}_m - \boldsymbol{\theta}^*)\|$ is not monotonous in m.

Hint: Consider e.g. $\widetilde{\boldsymbol{\theta}}_m$ being the projector on the first m coordinates and $W\boldsymbol{\theta} = \sum_j \theta_j$. Take $\boldsymbol{\theta}^*$ by alternating blocks of 1's and -1's with equal length.

Exercise 0.41. Let $\mathbb{E}\boldsymbol{\varepsilon} = 0$, $\operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$, and $\widetilde{\boldsymbol{\theta}}_m = \mathcal{S}_m \boldsymbol{Y}$. Then

$$\mathbb{E}\bigl\|W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}})\bigr\|^2 = \|W(\boldsymbol{\theta}_m^* - \boldsymbol{\theta}^*)\|^2 + \sigma^2 \operatorname{tr}\bigl\{W(\mathcal{S} - \mathcal{S}_m)(\mathcal{S} - \mathcal{S}_m)^\top W^\top\bigr\}.$$

Derive a similar expansion for an inhomogeneous noise $\boldsymbol{\varepsilon}$ with $\operatorname{Var}(\boldsymbol{\varepsilon}) = \Sigma$.

Exercise 0.42. Let $\operatorname{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$ and $\mathcal{S}_m = \Pi_m$ is a projector on the subspace spanned by $\theta_1, \ldots \theta_m$.

1. Check that

$$W\mathcal{S}\mathcal{S}_m^\top W^\top = W\mathcal{S}_m\mathcal{S}_m^\top W^\top = W\mathcal{S}_m W^\top.$$

2. Compute $\widetilde{\mathcal{R}}_m = \mathbb{E}\|W(\widetilde{\boldsymbol{\theta}}_m - \boldsymbol{\theta}^*)\|^2$ if W projects $\boldsymbol{\theta}$ onto the first q components.

20.01

Exercise 0.43. Let $D^2(\boldsymbol{\theta}) = -\nabla^2 I\!E L(\boldsymbol{\theta})$ and

$$\|I_p - D^{-1} D^2(\boldsymbol{\theta}) D^{-1}\|_\infty \, \mathtt{r} \le \delta(\mathtt{r})\mathtt{r}, \quad \forall \boldsymbol{\theta} \in \Theta_0(\mathtt{r}).$$

Show that for any $\boldsymbol{\theta}, \boldsymbol{\theta}^\circ \in \Theta_0(\mathtt{r})$

$$\bigl| D^{-1}\bigl\{ \nabla I\!E L(\boldsymbol{\theta}) - \nabla I\!E L(\boldsymbol{\theta}^\circ) \bigr\} - D(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \bigr| \le 2\mathtt{r}\delta(\mathtt{r})$$

and also

$$\begin{aligned} \Bigl| I\!E L(\boldsymbol{\theta}) - I\!E L(\boldsymbol{\theta}^*) - \bigl\| D(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \bigr\|^2/2 \Bigr| &\le \mathtt{r}^2 \delta(\mathtt{r})/2, \\ \Bigl| I\!E L(\boldsymbol{\theta}) - I\!E L(\boldsymbol{\theta}^\circ) - (\boldsymbol{\theta} - \boldsymbol{\theta}^\circ)^\top \nabla I\!E L(\boldsymbol{\theta}^\circ) - \bigl\| D(\boldsymbol{\theta} - \boldsymbol{\theta}^\circ) \bigr\|^2/2 \Bigr| &\le \mathtt{r}^2 \delta(\mathtt{r}). \end{aligned}$$

Exercise 0.44. Consider linear projection estimates $\widetilde{\boldsymbol{\theta}}_m$ in linear regression $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$. Check for each pair $m > m^\circ$ the identity

$$W\bigl(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^\circ}\bigr) = \boldsymbol{b}_{m,m^\circ} + \boldsymbol{\xi}_{m,m^\circ}$$

with

$$\begin{aligned} \boldsymbol{b}_{m,m^\circ} &\stackrel{\text{def}}{=} W(\boldsymbol{\theta}_m^* - \boldsymbol{\theta}_{m^\circ}^*), \\ \boldsymbol{\xi}_{m,m^\circ} &\stackrel{\text{def}}{=} W\bigl(D_m^{-2} - D_{m^\circ}^{-2}\bigr)\nabla. \end{aligned}$$

Exercise 0.45. Let

$$\delta_\Psi \stackrel{\text{def}}{=} \|V^{-1}\Sigma^{1/2}\Psi\|_\infty = \max_{i=1,\ldots,n} \|V^{-1}\Psi_i\|\sigma_i \, . \tag{0.7}$$

Consider the case of a regular design, when all the Ψ_i 's belongs to a compact subset $\mathcal{X}$ of $I\!R^p$, the matrix

$$d_\Psi^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \Psi_i \Psi_i^\top$$

is non-degenerate, and the value

$$a_\Psi \stackrel{\text{def}}{=} \max_{i=1,\ldots,n} \|d_\Psi^{-1}\Psi_i\|$$

is finite. Moreover, let the ratio $\lambda_{\max}(\Sigma)/\lambda_{\min}(\Sigma)$ of the maximal and minimal values of σ_i^2 's be bounded by a_Σ^2, that is, for all $i, j \le n$

$$\sigma_i/\sigma_j \le a_\Sigma \, ,$$

Check that δ_Ψ from (0.7) fulfills

$$\delta_\Psi \le a_\Psi \, a_\Sigma \, n^{-1/2}.$$

Exercise 0.46. Define

$$\boldsymbol{U}^\top \stackrel{\text{def}}{=} V^{-1}\Psi\Sigma^{1/2}.$$

Check that the columns $\boldsymbol{u}_j \in I\!R^n$ of the matrix $\boldsymbol{U}$ are orthonormal:

$$\boldsymbol{u}_i^\top \boldsymbol{u}_j = 1\!\!I(i = j).$$

Suppose that

$$\max_{i=1,\ldots,n} \|V^{-1}\Psi_i\|\sigma_i \le \delta_\Psi$$

Then for any unit vector $\boldsymbol{\gamma} \in I\!R^p$, the vector $\boldsymbol{u} = \boldsymbol{U}\boldsymbol{\gamma} \in I\!R^n$ fulfills $\|\boldsymbol{u}\| = 1$ and

$$\|\boldsymbol{u}\|_\infty \le \delta_\Psi \, .$$

Hint: suppose that all $\sigma_i \equiv 1$. By the Cauchy-Schwartz inequality, each component u_i of $\boldsymbol{u}$ satisfies

$$u_i = \boldsymbol{\gamma}^\top V^{-1}\Psi_i \le \|\boldsymbol{\gamma}\| \, \|V^{-1}\Psi_i\|.$$

27.01

Exercise 0.47. Write the sieve MLE $\widetilde{\boldsymbol{\theta}}_m$ for the generalized linear regression $Y_i \sim P_{\Psi_i^\top \boldsymbol{\theta}} \in \mathcal{P}$ in

- Exponential regression with the canonical parameter $\boldsymbol{\upsilon} = 1/\lambda$, where λ is the Poisson intensity parameter;
- logit (Bernoulli canonical) model

Describe in each case the Fisher matrices D_m, the score ∇_m, and the stochastic term $\boldsymbol{\xi}_{m,m^\circ}$ in expansion

$$\boldsymbol{\xi}_{m,m^\circ} \stackrel{\text{def}}{=} W\bigl(D_m^{-2} - D_{m^\circ}^{-2}\bigr)\nabla = W\bigl(D_m^{-2} - D_{m^\circ}^{-2}\bigr)\nabla_m .$$

Exercise 0.48. Let $w_i^\flat \sim \mathcal{N}(1,1)$ be Gaussian bootstrap multiplier for the linear Gaussian model $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta} + \boldsymbol{\varepsilon}$. For linear sieve estimates $\widetilde{\boldsymbol{\theta}}_m$ with

$$\widetilde{\boldsymbol{\theta}}_m = \bigl(\Psi_m \Psi_m^\top\bigr)^{-1} \Psi_m \mathcal{W}^\flat \boldsymbol{Y}$$

check whether

$$\begin{aligned}
I\!\!E^\flat W(\widetilde{\boldsymbol{\theta}}_m^\flat - \widetilde{\boldsymbol{\theta}}_{m^\circ}^\flat) &= W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^\circ}), \\
I\!\!E^\flat \|W(\widetilde{\boldsymbol{\theta}}_m^\flat - \widetilde{\boldsymbol{\theta}}_{m^\circ}^\flat)\| &= \|W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^\circ})\|, \\
I\!\!E^\flat \|W(\widetilde{\boldsymbol{\theta}}_m^\flat - \widetilde{\boldsymbol{\theta}}_{m^\circ}^\flat)\|^2 &= \|W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^\circ})\|^2.
\end{aligned}$$

If not equal, compare $I\!\!E^\flat \|W(\widetilde{\boldsymbol{\theta}}_m^\flat - \widetilde{\boldsymbol{\theta}}_{m^\circ}^\flat)\|^2$ and $\|W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^\circ})\|^2$.

Exercise 0.49. Consider the linear Gaussian model $\boldsymbol{Y} = \Psi^\top \boldsymbol{\theta} + \boldsymbol{\varepsilon}$ and let $w_i^\flat \sim \mathcal{N}(1,1)$ be Gaussian bootstrap multiplier. Let also $\mathcal{E}^\flat \boldsymbol{\varepsilon}$ be the vector with the entries $w_i^\flat \varepsilon_i$.

- Check that

$$\operatorname{Var}^\flat(\mathcal{E}^\flat \boldsymbol{\varepsilon}) = \operatorname{diag}(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}),$$

  where $\operatorname{diag}(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon}) = \operatorname{diag}(\varepsilon_1^2, \ldots, \varepsilon_n^2)$ is the diagonal matrix with the entries ε_i^2.

- Check the formula

$$\operatorname{Var}^\flat(\Psi \mathcal{E}^\flat \boldsymbol{\varepsilon}) = I\!\!E^\flat\bigl\{\Psi \mathcal{E}^\flat \boldsymbol{\varepsilon} (\Psi \mathcal{E}^\flat \boldsymbol{\varepsilon})^\top\bigr\} = \Psi \operatorname{Var}(\mathcal{E}^\flat \boldsymbol{\varepsilon})\, \Psi^\top = \Psi \operatorname{diag}(\boldsymbol{\varepsilon} \cdot \boldsymbol{\varepsilon})\, \Psi^\top.$$

Exercise 0.50. Consider the linear Gaussian regression model $\boldsymbol{Y} = \Psi^{\top}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ and linear bootstrap estimates $\widetilde{\boldsymbol{\theta}}_m^{\flat}(\boldsymbol{w}^{\flat}) = \bigl(\Psi_m \Psi_m^{\top}\bigr)^{-1} \Psi_m \mathcal{W}^{\flat} \boldsymbol{Y}$. Compute the bootstrap mean and bootstrap variance of the sum $\frac{1}{B} \sum_{\boldsymbol{w}^{\flat}} \widetilde{\boldsymbol{\theta}}_m^{\flat}(\boldsymbol{w}^{\flat})$:

$$\mathbb{E}^{\flat}\biggl\{ \frac{1}{B} \sum_{\boldsymbol{w}^{\flat}} \widetilde{\boldsymbol{\theta}}_m^{\flat}(\boldsymbol{w}^{\flat}) \biggr\} = ?,$$

$$\operatorname{Var}^{\flat}\biggl\{ \frac{1}{B} \sum_{\boldsymbol{w}^{\flat}} \widetilde{\boldsymbol{\theta}}_m^{\flat}(\boldsymbol{w}^{\flat}) \biggr\} = ?$$

Exercise 0.51. For the linear Gaussian regression model and linear estimates $\widetilde{\boldsymbol{\theta}}_m^{\flat}(\boldsymbol{w}^{\flat}) = \bigl(\Psi_m \Psi_m^{\top}\bigr)^{-1} \Psi_m \mathcal{W}^{\flat} \boldsymbol{Y}$, compute the value $\mathtt{p}_{m,m^{\circ}}^{\flat} = \mathbb{E}^{\flat} \bigl\| \boldsymbol{\xi}_{m,m^{\circ}}^{\flat} \bigr\|^2$. Check whether the Monte-Carlo sum

$$\frac{1}{B} \sum_{\boldsymbol{w}^{\flat}} \bigl\| W(\widetilde{\boldsymbol{\theta}}_m^{\flat} - \widetilde{\boldsymbol{\theta}}_{m^{\circ}}^{\flat}) - W(\widetilde{\boldsymbol{\theta}}_m - \widetilde{\boldsymbol{\theta}}_{m^{\circ}}) \bigr\|^2$$

is an unbiased estimate of this value.

02.02

For any two measures $I\!P, I\!P_1$ on the same probability space, define the total variation distance as

$$\|I\!P - I\!P_1\|_{TV} \stackrel{\text{def}}{=} \sup_A |I\!P(A) - I\!P_1(A)| \tag{0.8}$$

where maximum is taken over all measurable subsets A.

Exercise 0.52. Let $I\!P, I\!P_1$ be dominated by $\boldsymbol{\mu}_0$ and $\rho(\omega) = dI\!P/d\boldsymbol{\mu}_0(\omega)$, $\rho_1(\omega) = dI\!P_1/d\boldsymbol{\mu}_0(\omega)$ be the corresponding density functions.

- Show that

$$\|I\!P - I\!P_1\|_{TV} \le \int |\rho(\omega) - \rho_1(\omega)| d\boldsymbol{\mu}_0(\omega)$$

- Show that

$$\int \bigl[\rho(\omega) - \rho_1(\omega)\bigr]_+ d\boldsymbol{\mu}_0(\omega) = \int \bigl[\rho_1(\omega) - \rho(\omega)\bigr]_+ d\boldsymbol{\mu}_0(\omega).$$

  Here $x_+ = \max\{x, 0\}$.

- Describe A for which the maximum in (0.8) is attained.

- Show that

$$\|I\!P - I\!P_1\|_{TV} \le \frac{1}{2} \int |\rho(\omega) - \rho_1(\omega)| d\boldsymbol{\mu}_0(\omega).$$

Hint: use that

$$I\!P(A) - I\!P_1(A) = \int \bigl[\rho(\omega) - \rho_1(\omega)\bigr] d\boldsymbol{\mu}_0(\omega),$$
$$\int \rho(\omega) d\boldsymbol{\mu}_0(\omega) = \int \rho_1(\omega) d\boldsymbol{\mu}_0(\omega) = 1.$$

Exercise 0.53. Let the distribution $\mathbb{Q}$ of the score $\nabla = \Psi\boldsymbol{\varepsilon}$ and the conditional distribution $\mathbb{Q}^\flat$ of the bootstrap score $\nabla^\flat = \Psi\mathcal{E}^\flat\boldsymbol{Y}$ given $\boldsymbol{Y}$ be related by

$$\bigl\|\mathbb{Q} - \mathbb{Q}^\flat\bigr\|_{TV} \le \delta$$

for some δ. Show that the same bounds applies to the distance between the distribution of the qMLE $\widetilde{\boldsymbol{\theta}}$ and of the bootstrap estimate $\widetilde{\boldsymbol{\theta}}^\flat = \bigl(\Psi\Psi^\top\bigr)^{-1}\Psi\mathcal{W}^\flat\boldsymbol{Y}$. Derive as a special case that

$$\sup_{z>0}\Bigl| I\!P\bigl(\|W(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\| \le z\bigr) - I\!P^\flat\bigl(\|W(\widetilde{\boldsymbol{\theta}}^\flat - \widetilde{\boldsymbol{\theta}}\| \le z\bigr)\Bigr| \le \delta.$$

Exercise 0.54. Let $\nabla^{\flat} \sim \mathcal{N}(0, \mathcal{V}^2)$ and $\nabla \sim \mathcal{N}(0, V^2)$, and

$$\mathcal{V}^2 \geq V^2 .$$

Show that

$$\sup_{z>0} \Bigl\{ \mathbb{P}\bigl(\| W\bigl(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\bigr) \| \leq z \bigr) - \mathbb{P}^{\flat}\bigl(\| W\bigl(\widetilde{\boldsymbol{\theta}}^{\flat} - \widetilde{\boldsymbol{\theta}} \| \leq z \bigr) \Bigr\} \geq 0.$$

Use that the variance of $\nabla^{\flat}$ is larger than the variance of ∇, the related probability of any centrally symmetric convex set A in $\mathbb{R}^p$ is larger under $\mathbb{P}$ than under $\mathbb{P}^{\flat}$.

Exercise 0.55. Let $\mathbb{P} = \mathcal{N}(0, V^2)$ and $\mathbb{P}^{\flat} = \mathcal{N}(0, \mathcal{V}^2)$ with $\mathcal{V}^2 = aV^2$ for some $a \geq 1$ and a positive symmetric $p \times p$ matrix V^2. Compute $\mathcal{K}(\mathbb{P}, \mathbb{P}^{\flat}) = \mathbb{E} \log\bigl(d\mathbb{P} / d\mathbb{P}^{\flat} \bigr)$ and $\mathcal{K}(\mathbb{P}^{\flat}, \mathbb{P})$.

Exercise 0.56. Let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be two standard normal vectors in $\mathbb{R}^p$. Define $\boldsymbol{\xi}_1 = \boldsymbol{\xi} + \sigma \boldsymbol{\eta}$. Compute $\mathcal{K}(\mathbb{P}_1, \mathbb{P})$, where $\mathbb{P}$ (rest. $\mathbb{P}_1$) means the law of $\boldsymbol{\xi}$ (rest. of $\boldsymbol{\xi}_1$).