Asymptotic Theory of Robustness
a short summary

Matthias Kohl

"Mathematical Statistics"

University of Bayreuth

Contents

1 Introduction

In this technical report we give a short summary of results contained in Chapters 2 - 5 of [Rieder, 1994] where we restrict our considerations to the estimation of a finite-dimensional parameter in the one sample i.i.d. case (i.e., no testing, no functionals). More precisely, we assume a parametric family

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A}) \tag{1.1}$$

of probability measures on some sample space $(\Omega, \mathcal{A})$, whose parameter space Θ is an open subset of some finite dimensional $\mathbb{R}^k$. Sections 2 - 5 contain a short abridge of some classical results of asymptotic statistics. For a more detailed introduction to this topics we also refer to Chapter 2 of [Bickel et al., 1998] and Chapters 6 - 9 of [van der Vaart, 1998], respectively. In the infinitesimal robust setup introduced in Section 6, the family $\mathcal{P}$ will serve as ideal center model and at least under the null hypothesis $P_\theta \in \mathcal{P}$ the observations $y_1, \ldots, y_n$ at time $n \in \mathbb{N}$ are assumed to be i.i.d.. Finally, in Section 7 we give the solutions (i.e., optimal influence curves) to the optimization problems motivated in Subsection 7.1. For the derivation of optimal influence curves confer also [Hampel, 1968] and [Hampel et al., 1986].

2 L_2 differentiability

To avoid domination assumptions in the definition of L_2 differentiability, we employ the following square root calculus that was introduced by Le Cam. The following definition is taken from [Rieder, 1994]; for more details confer Subsection 2.3.1 of [Rieder, 1994].

Definition 2.1 *For any measurable space $(\Omega, \mathcal{A})$ and $k \in \mathbb{N}$ we define the following real Hilbert space that includes the ordinary $L_2^k(P)$*

$$\mathcal{L}_2^k(\mathcal{A}) = \{\xi\sqrt{dP} \,|\, \sigma \in L_2^k(P), P \in \mathcal{M}_b(\mathcal{A})\} \tag{2.1}$$

On this space, an equivalence relation is given by

$$\xi\sqrt{dP} \equiv \eta\sqrt{dQ} \qquad \Longleftrightarrow \qquad \int |\xi\sqrt{p} - \eta\sqrt{q}\,|^2 d\mu = 0 \tag{2.2}$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^k$ and $\mu \in \mathcal{M}_b(\mathcal{A})$ may be any measure, depending on P and Q, so that $dP = p\,d\mu$, $dQ = q\,d\mu$. We define linear combinations *with real coefficients and a scalar product by*

$$\alpha\xi\sqrt{dP} + \beta\eta\sqrt{dQ} = (\alpha\xi\sqrt{p} + \beta\eta\sqrt{q}\,)\sqrt{d\mu} \tag{2.3}$$

$$\langle \xi\sqrt{dP} \,|\, \eta\sqrt{dQ}\,\rangle = \int \xi^\tau \eta \sqrt{pq}\, d\mu \tag{2.4}$$

We fix some $\theta \in \Theta$ and define L_2 differentiability of the family $\mathcal{P}$ at θ using this square root calculus; confer Definition 2.3.6 of [Rieder, 1994]. Here E_θ denotes expectation taken under P_θ.

Definition 2.2 *Model $\mathcal{P}$ is called L_2 differentiable at θ if there exists some function $\Lambda_\theta \in L_2^k(P_\theta)$ such that, as $t \to 0$,*

$$\|\sqrt{dP_{\theta+t}} - \sqrt{dP_\theta}\,(1 + \tfrac{1}{2}t^\tau\Lambda_\theta)\|_{\mathcal{L}_2^k} = o(|t|) \tag{2.5}$$

and

$$\mathcal{I}_\theta = \mathrm{E}_\theta\, \Lambda_\theta\Lambda_\theta^\tau \succ 0 \tag{2.6}$$

The function Λ_θ is called the L_2 derivative and the $k\times k$ matrix $\mathcal{I}_\theta$ Fisher Information of $\mathcal{P}$ at θ.

Remark 2.3 A concise definition of L_2 differentiability for arrays of probability measures on general sample spaces may be found in Section 2.3 of [Rieder, 1994]. ////

We now consider a parameter sequence (θ_n) about θ of the form

$$\theta_n = \theta + \frac{t_n}{\sqrt{n}} \qquad\qquad t_n \to t \in \mathbb{R}^k \tag{2.7}$$

Corresponding to this parametric alternatives (θ_n) two sequences of product measures are defined on the n-fold product measurable space $(\Omega^n, \mathcal{A}^n)$

$$P_\theta^n = \bigotimes_{i=1}^n P_\theta \qquad\qquad P_{\theta_n}^n = \bigotimes_{i=1}^n P_{\theta_n} \tag{2.8}$$

Theorem 2.4 *If $\mathcal{P}$ is L_2 differentiable at θ, its L_2 derivative Λ_θ is uniquely determined in $L_2^k(P_\theta)$. Moreover,*

$$\mathrm{E}_\theta\, \Lambda_\theta = 0 \tag{2.9}$$

and the alternatives given by (2.7) and (2.8) have the log likelihood expansion

$$\log \frac{dP_{\theta_n}^n}{dP_\theta^n} = \frac{t^\tau}{\sqrt{n}} \sum_{i=1}^n \Lambda_\theta(y_i) - \frac{1}{2} t^\tau \mathcal{I}_\theta t + o_{P_\theta^n}(n^0) \tag{2.10}$$

where

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Lambda_\theta(y_i)\right)(P_\theta^n) \xrightarrow{\mathrm{w}} \mathcal{N}(0, \mathcal{I}_\theta) \tag{2.11}$$

Proof: This is a special case of Theorem 2.3.7 in [Rieder, 1994]. ////

3 Local Asymptotic Normality

We first state a result of asymptotic statistics that is known as *Le Cam's third lemma*.

Theorem 3.1 *Let $P_n, Q_n \in \mathcal{M}_1(\mathcal{A}_n)$ be two sequences of probabilities with log likelihoods $L_n \in \log \frac{dQ_n}{dP_n}$, and S_n a sequence of statistics on $(\Omega_n, \mathcal{A}_n)$ taking values in some finite-dimensional $(\bar{\mathbb{R}}^p, \bar{\mathbb{B}}^p)$ such that for $a, c \in \mathbb{R}^p$, $\sigma \in [0, \infty)$ and $C \in \mathbb{R}^{p \times p}$,*

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix}(P_n) \xrightarrow{\mathrm{w}} \mathcal{N}\left(\begin{pmatrix} a \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} C & c \\ c^\tau & \sigma^2 \end{pmatrix}\right) \tag{3.1}$$

then

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix}(Q_n) \xrightarrow{\mathrm{w}} \mathcal{N}\left(\begin{pmatrix} a + c \\ \sigma^2/2 \end{pmatrix}, \begin{pmatrix} C & c \\ c^\tau & \sigma^2 \end{pmatrix}\right) \tag{3.2}$$

$$\tag{3.3}$$

PROOF: [Rieder, 1994], Corollary 2.2.6. ////

The following definition corresponds to Definition 2.2.9 of [Rieder, 1994].

Definition 3.2 *A sequence $(\mathcal{Q}_n)$ of statistical models on sample spaces $(\Omega_n, \mathcal{A}_n)$,*

$$\mathcal{Q}_n = \{Q_{n,t} \,|\, t \in \Theta_n\} \subset \mathcal{M}_1(\mathcal{A}_n) \tag{3.4}$$

with the same finite-dimensional parameter space $\Theta_n = \mathbb{R}^k$ (or at least $\Theta_n \uparrow \mathbb{R}^k$) is called asymptotically normal, *if there exists a sequence of random variables $Z_n\colon (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}^k, \mathbb{B}^k)$ that are asymptotically normal,*

$$Z_n(Q_{n,0}) \xrightarrow{\mathrm{w}} \mathcal{N}(0, C) \tag{3.5}$$

with positive definite covariance $C \in \mathbb{R}^{k \times k}$, and such that for all $t \in \mathbb{R}^k$ the log likelihoods $L_{n,t} \in \log \frac{dQ_{n,t}}{dQ_{n,0}}$ have the approximation

$$L_{n,t} = t^\tau Z_n - \frac{1}{2} t^\tau C t + o_{Q_{n,0}}(n^0) \tag{3.6}$$

The sequence $Z = (Z_n)$ is called the asymptotically sufficient statistic *and C the* asymptotic covariance *of the* asymptotically normal models $(\mathcal{Q}_n)$.

We now state Remark 2.2.10 of [Rieder, 1994] where we add a part (c) and (d).

Remark 3.3 **(a)** The covariance C is uniquely defined by (3.5) and (3.6). And (3.6) implies that another sequence of statistics $W = (W_n)$ is asymptotically sufficient iff $W_n = Z_n + o_{Q_{n,0}}(n^0)$.

(b) Neglecting the approximation, the terminology of *asymptotically sufficient* may be justified in regard to Neyman's criterion; confer Proposition C.1.1 of [Rieder, 1994]. One speaks of *local asymptotic normality* if, as in section 2, asymptotic normality depends on suitable local reparametrizations.

(c) The notion of *local asymptotic normality* – in short, LAN – was introduced by [Le Cam, 1960].

(d) The sequence of statistical models $\mathcal{Q}_n = \{Q_{n,t} \,|\, Q_{n,t} = P^n_{\theta_n}\}$ given by the alternatives (2.7) and (2.8) is LAN with asymptotically sufficient statistic $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Lambda_\theta(y_i)$ and asymptotic covariance $C = \mathcal{I}_\theta$. ////

4 Convolution Representation and Asymptotic Minimax Bound

In this section we present the convolution and the asymptotic minimax theorems in the parametric case; confer Theorems 3.2.3, 3.3.8 of [Rieder, 1994]. These two mathematical results of asymptotic statistics are mainly due to Le Cam and Hájek.
Assume a sequence of statistical models $(\mathcal{Q}_n)$ on sample spaces $(\Omega_n, \mathcal{A}_n)$,

$$\mathcal{Q}_n = \{Q_{n,t} \,|\, t \in \Theta_n\} \subset \mathcal{M}_1(\mathcal{A}_n) \tag{4.1}$$

with the same finite-dimensional parameter space $\Theta_n = \mathbb{R}^k$ (or $\Theta_n \uparrow \mathbb{R}^k$). The parameter of interest is Dt for some $p \times k$-matrix D of full rank $p \le k$. Moreover we consider asymptotic estimators

$$S = (S_n) \qquad S_n \colon (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}^p, \mathbb{B}^p) \tag{4.2}$$

The following definition corresponds to Definition 3.2.2 of [Rieder, 1994].

Definition 4.1 *An asymptotic estimator S is called* regular *for the parameter transform D, with limit law $M \in \mathcal{M}_1(\mathbb{B}^p)$, if for all $t \in \mathbb{R}^k$,*

$$(S_n - Dt)(Q_{n,t}) \xrightarrow{\mathrm{w}} M \tag{4.3}$$

that is, $S_n(Q_{n,t}) \xrightarrow{\mathrm{w}} M * \mathrm{I}_{Dt}$ as $n \to \infty$, for every $t \in \mathbb{R}^k$.

Remark 4.2 For a motivation of this regularity assumption we refer to Example 3.2.1 of [Rieder, 1994]. The Hodges estimator introduced there is asymptotically normal but superefficient. However it is not regular in the sense of Definitions 4.1. Moreover, in the light of the asymptotic minimax theorem (Theorem 4.5), the Hodges estimator has maximal estimator risk; confer [Rieder, 1994], Example 3.3.10. ////

We now may state the convolution theorem.

Theorem 4.3 *Assume models $(\mathcal{Q}_n)$ that are asymptotically normal with asymptotic covariance $C \succ 0$ and asymptotically sufficient statistic $Z = (Z_n)$. Let $D \in \mathbb{R}^{p \times k}$ be a matrix of rank $p \le k$. Let the asymptotic estimator S be regular for D with limit law M. Then there exists a probability $M_0 \in \mathcal{M}_1(\mathbb{B}^p)$ such that*

$$M = M_0 * \mathcal{N}(0, \Gamma) \qquad \Gamma = DC^{-1}D^\tau \tag{4.4}$$

and

$$(S_n - DC^{-1}Z_n)(Q_{n,0}) \xrightarrow{\mathrm{w}} M_0 \tag{4.5}$$

An asymptotic estimator S^* is regular for D and achieves limit law $M^* = \mathcal{N}(0,\Gamma)$ iff

$$S_n^* = DC^{-1}Z_n + o_{Q_{n,0}}(n^0) \tag{4.6}$$

PROOF: Three variants of the proof are given in [Rieder, 1994], Theorem 3.2.3. ////

For the specification of the asymptotic minimax theorem we need the definition of the set L of loss functions; confer pp. 78, 81 of [Rieder, 1994].

Definition 4.4 *Let* L *be the set of all Borel measurable functions $\ell\colon \bar{\mathbb{R}}^p \to [0,\infty]$ that are*

(a) symmetric subconvex *on $\mathbb{R}^p$; that is, for all $z \in \mathbb{R}^p$ and all $c \in [0,\infty]$,*

$$\ell(z) = \ell(-z) \qquad \{z \in \mathbb{R}^p \,|\, \ell(z) \le c\} \text{ is convex} \tag{4.7}$$

(b) upper semicontinuous at infinity*; that is, for every sequence $z_n \in \mathbb{R}^p$ with $z_n \to z \in \bar{\mathbb{R}}^p \setminus \mathbb{R}^p$,*

$$\limsup_{n\to\infty} \ell(z_n) \le \ell(z) \tag{4.8}$$

This functions $\ell \in$ L will be called loss functions*. If there is an increasing function $v\colon [0,\infty] \to [0,\infty]$ and a symmetric positive definite matrix $A \in \mathbb{R}^{p\times p}$, then a loss function of type,*

$$\ell(z) = \begin{cases} v(z^\tau A z) & \text{if } |z| < \infty \\ v(\infty) & \text{if } |z| = \infty \end{cases} \tag{4.9}$$

will be called monotone quadratic.

For part (a) of the asymptotic minimax theorem we assume Θ_n open. Moreover asymptotic estimators with extended values can be allowed; i.e.,

$$S = (S_n) \qquad S_n\colon (\Omega_n, \mathcal{A}_n) \to (\bar{\mathbb{R}}^p, \bar{\mathbb{B}}^p) \tag{4.10}$$

Theorem 4.5 *Assume models $(\mathcal{Q}_n)$ that are asymptotically normal with asymptotic covariance $C \succ 0$. Let $D \in \mathbb{R}^{p\times k}$ be a matrix of rank $p \le k$. Put*

$$\rho_0 = \int \ell \, d\mathcal{N}(0,\Gamma) \qquad \Gamma = DC^{-1}D^\tau \tag{4.11}$$

for any Borel measurable function $\ell\colon \mathbb{R}^p \to [0,\infty]$.

(a) *Then, if $\ell \in$ L and ℓ is lower semicontinuous on $\bar{\mathbb{R}}^p$,*

$$\lim_{b\to\infty}\lim_{c\to\infty}\liminf_{n\to\infty}\inf_S \sup_{|t|\le c} \int b \wedge \ell(S_n - Dt)\, dQ_{n,t} \ge \rho_0 \tag{4.12}$$

(b) *Suppose $\ell\colon \mathbb{R}^p \to [0,\infty]$ is continuous a.e. λ^p and the asymptotic estimator S^* is asymptotically normal for every $c \in (0,\infty)$, uniformly in $|t| \le c$,*

$$(S_n^* - Dt)(Q_{n,t}) \xrightarrow{w} \mathcal{N}\,(0,\Gamma) \tag{4.13}$$

Then for all $c \in (0,\infty)$,

$$\lim_{b\to\infty} \lim_{n\to\infty} \sup_{|t|\le c} \int b \wedge \ell(S_n^* - Dt)\, dQ_{n,t} = \rho_0 \tag{4.14}$$

and necessarily

$$S_n^* = DC^{-1}Z_n + o_{Q_{n,0}}(n^0) \tag{4.15}$$

Proof: [Rieder, 1994], Theorem 3.3.8 and Remark 3.3.9 (d). ////

5 Asymptotically Linear Estimators

In this section we define influence curves (ICs) respectively, partial ICs and asymptotically linear estimators (ALEs). Moreover we derive the Cramér-Rao bound in the smooth parametric i.i.d. case and restricted to the class of ALEs.

5.1 Definitions

Often ICs are introduced as Gâteaux derivatives of statistical functionals; confer Section 2.5 of [Huber, 1981] and Section 2.1 of [Hampel et al., 1986], respectively. But most proofs of asymptotic normality in the i.i.d. case head for an estimator expansion, in which ICs canonically occur as summands; confer M, L, R, S and MD (minimum distance) estimates. The following definition corresponds to Definition 4.2.10 of [Rieder, 1994].

Definition 5.1 *Suppose $\mathcal{P}$ is L_2 differentiable at θ, and assume some matrix $D \in \mathbb{R}^{p\times k}$ of full rank $p \le k$. Let $\alpha = 2, \infty$, respectively.*

(a) *Then the set $\Psi_2(\theta)$ of all* square integrable *and the subset $\Psi_\infty(\theta)$ of all* bounded influence curves *at P_θ , respectively, are*

$$\Psi_\alpha(\theta) = \big\{\psi_\theta \in L_2^k(P_\theta) \,|\, \mathrm{E}_\theta\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k\big\} \tag{5.1}$$

(b) *The set $\Psi_2^D(\theta)$ of all* square integrable *and the subset $\Psi_\infty^D(\theta)$ of all* bounded, partial influence curves *at P_θ, respectively, are*

$$\Psi_\alpha^D(\theta) = \big\{\psi_\theta \in L_2^p(P_\theta) \,|\, \mathrm{E}_\theta\, \psi_\theta = 0,\ \mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = D\big\} \tag{5.2}$$

In this context we repeat Remark 4.2.11 of [Rieder, 1994] where we omit part (d) about L_1 differentiability, note part (e) without the proof and add *time series models* to part (f).

Remark 5.2 **(a)** The attribute *square integrable* will usually be omitted.

(b) The *classical scores* and the *classical partial scores*,

$$\psi_{h,\theta} = \mathcal{I}_\theta^{-1}\Lambda_\theta \in \Psi_2(\theta) \tag{5.3}$$
$$\eta_{h,\theta} = D\psi_{h,\theta} = D\mathcal{I}_\theta^{-1}\Lambda_\theta \in \Psi_2^D(\theta) \tag{5.4}$$

are always ICs, respectively, partial ICs, at P_θ.

(c) The definition of $\Psi_2(\theta)$ and $\Psi_\infty(\theta)$ requires $\mathcal{I}_\theta \succ 0$, and Λ_θ nondegenerate in the sense that, for all $t \in \mathbb{R}^k$,

$$t^\tau \Lambda_\theta = 0 \quad \text{a.e. } P_\theta \qquad \Longrightarrow \qquad t = 0 \tag{5.5}$$

(d) [...]

(e) [...] $\Psi_\alpha^D(\theta) = \big\{D\psi_\theta \,|\, \psi_\theta \in \Psi_\alpha(\theta)\big\}$ [...]

(f) Of course, $\Psi_\alpha^D(\theta) = \Psi_\alpha(\theta)$ for $D = \mathbb{I}_k$. Partial ICs with general D occur when there are nuisance components. In robust regression – *respectively, time series models* –, moreover, conditionally centered (partial) ICs will occur.

(g) $\Psi_\alpha^D(\theta)$ are closed convex subsets of $L_\alpha^p(P_\theta)$; $\alpha = 2, \infty$. ////

Next we give the definition of asymptotically linear estimators (ALEs); confer Definition 4.2.16 of [Rieder, 1994].

Definition 5.3 *An asymptotic estimator*

$$S = (S_n) \qquad S_n : (\Omega^n, \mathcal{A}^n) \to (\mathbb{R}^k, \mathbb{B}^k) \tag{5.6}$$

is called asymptotically linear at P_θ *if there is an IC* $\psi_\theta \in \Psi_2(\theta)$ *such that*

$$R_n = \sqrt{n}\,(S_n - \theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_\theta(y_i) + o_{P_\theta^n}(n^0) \tag{5.7}$$

We call $R = (R_n)$ standardization, *and* ψ_θ *the* IC, *of* S *at* P_θ.

We now state Remark 4.2.17 of [Rieder, 1994] where we omit part (c) on L_1 differentiability and part (f) on the nonparametric convolution and asymptotic minimax theorems.

Remark 5.4 **(a)** The expansion (5.7) determines the IC ψ_θ uniquely, because $\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta(y_i)$ with $\eta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta\,\eta = 0$, can tend to 0 in P_θ^n probability only if $\mathrm{E}_\theta\,|\eta|^2 = 0$; that is, $\eta = 0$ a.e. P_θ.

(b) If S is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$, then

$$\sqrt{n}\,(S_n - \theta)(P_\theta^n) \xrightarrow{\;\mathrm{w}\;} \mathcal{N}\left(0, \mathrm{Cov}_\theta(\psi_\theta)\right) \tag{5.8}$$

because of $\psi_\theta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta\,\psi_\theta = 0$, and the Lindeberg-Lévy theorem. The third condition $\mathrm{E}_\theta\,\psi_\theta\Lambda_\theta^\tau = \mathbb{I}_k$, as already noted in the remarks of [Rieder, 1980] (p. 108), is equivalent to the locally uniform extension of this asymptotic normality; see Lemma 5.5 below.

(c) [...]
(d) Extending general M estimates, the class of ALEs has in the case $k = 1$ been introduced by [Rieder, 1980]. [Bickel, 1981] defined the related notion CULAN, employing however compact subsets of Θ instead of compacts in the local parameter space.
(e) The class of ALEs contains the common asymptotically normal M, L, R and MD (minimum distance) estimates; confer chapters 1 and 6 of [Rieder, 1994]. In fact, most proofs of asymptotic normality in the i.i.d. case end up with an extension (5.7); the corresponding conditions need to be verified only under the ideal model.
(f) [...]
(g) The previous robustness theories of [Huber, 1964], [Hampel, 1974], [Rieder, 1980] and [Bickel, 1981] have been formulated but for ALEs or, even more specialized, for M estimates. ////

The following lemma corresponds to Lemma 4.2.18 of [Rieder, 1994]. It is a consequence of Theorem 2.4 together with Slutzky's lemma, the Cramér-Wold devive and Le Cam's third lemma (Theorem 3.1).

Lemma 5.5 *Let the ALE S have the asymptotic expansion (5.7) involving some function $\psi_\theta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta\, \psi_\theta = 0$. Then*

$$\mathrm{E}_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k \tag{5.9}$$

holds iff

$$\sqrt{n}\,(S_n - \theta)(P_{\theta + t_n/\sqrt{n}}) \xrightarrow{\;\mathrm{w}\;} \mathcal{N}\,(t, \mathrm{Cov}_\theta(\psi_\theta)) \tag{5.10}$$

for all convergent sequences $t_n \to t$ in $\mathbb{R}^k$.

5.2 Cramér-Rao Bound

In this subsection we show that in the parametric setup, and restricted to the class of ALEs, the convolution theorem (Theorem 4.3) and the local asymptotic minimax theorem (Theorem 4.5) coincide with the Cramér-Rao bound. Therefore we first specialize these two theorems to the parametric context where one wants to estimate the parameter t of the product models

$$\mathcal{Q}_n = \left\{ P^n_{\theta + t/\sqrt{n}} \,|\, t \in \mathbb{R}^k \right\} \subset \mathcal{M}_1(\mathcal{A}) \tag{5.11}$$

The following proposition corresponds to Proposition 4.2.19 of [Rieder, 1994].

Proposition 5.6 **(a)** *Let an asymptotic estimator R be regular for t with limit law $M \in \mathcal{M}_1(\mathbb{B}^k)$. Then there is a probability $M_0 \in \mathcal{M}_1(\mathbb{B}^k)$ such that*

$$M = M_0 * \mathcal{N}\,(0, \mathcal{I}_\theta^{-1}) \tag{5.12}$$

A regular estimator R^* achieves the limit law $M^* = \mathcal{N}\,(0, \mathcal{I}_\theta^{-1})$ iff R^* is the standardization of an estimator S^* that is asymptotically linear at P_θ with IC $\psi_{h,\theta}$.

(b) *Let the loss function $\ell \in \mathrm{L}$ be lower semicontinuous on $\bar{\mathbb{R}}^k$. Then*

$$\lim_{b\to\infty}\lim_{c\to\infty}\liminf_{n\to\infty}\inf_{R}\sup_{|t|\le c}\int b\wedge \ell(R_n-t)\,dP^n_{\theta+t/\sqrt{n}} \ge \rho_0 \tag{5.13}$$

where

$$\rho_0 = \int \ell\, d\mathcal{N}(0,\mathcal{I}_\theta^{-1}) \tag{5.14}$$

If the function $\ell\colon \mathbb{R}^k \to [0,\infty]$ is continuous a.e. λ^k, and the estimator S^* is asymptotically linear at P_θ with IC $\psi_{h,\theta}$, then

$$\lim_{b\to\infty}\lim_{c\to\infty}\lim_{n\to\infty}\sup_{|t|\le c}\int b\wedge \ell(\sqrt{n}\,(S_n^*-\theta)-t)\,dP^n_{\theta+t/\sqrt{n}} = \rho_0 \tag{5.15}$$

Proof: By the LAN property of $\mathcal{Q}_n$ and Lemma 5.5, which ensures the uniformity on t-compacts of weak convergence needed for (5.15), we may apply Theorems 4.3 and 4.5. ////

Now we adapt this results to ALEs which are regular in the sense of Definition 4.1 and thus obtain the Cramér-Rao bound.

Proposition 5.7 *Consider an estimator $S = (S_n)$ that is asymptotically linear at P_θ with IC $\rho_\theta \in \Psi_2(\theta)$.*

(a) *Then its standardization R is regular with normal limit law*

$$\mathcal{N}(0,\mathrm{Cov}_\theta(\rho_\theta)) = \mathcal{N}(0,\mathrm{Cov}_\theta(\rho_\theta)-\mathcal{I}_\theta^{-1}) * \mathcal{N}(0,\mathcal{I}_\theta^{-1}) \tag{5.16}$$

(b) *Assume a loss function $\ell \in \mathrm{L}$ that is continuous a.e. λ^k. Then*

$$\begin{aligned}\lim_{b\to\infty}\lim_{c\to\infty}\lim_{n\to\infty}\sup_{|t|\le c}&\int b\wedge \ell(\sqrt{n}\,(S_n-\theta)-t)\,dP_{\theta+t/\sqrt{n}}\\ &= \int \ell\, d\mathcal{N}(0,\mathrm{Cov}_\theta(\rho_\theta)) \ge \int \ell\, d\mathcal{N}(0,\mathcal{I}_\theta^{-1})\end{aligned} \tag{5.17}$$

The lower bound is achieved by $\rho_\theta = \psi_{h,\theta}$. If ℓ is monotone quadratic and not constant a.e. λ^k, the lower bound can be achieved only by $\rho_\theta = \psi_{h,\theta}$.

Proof: [Rieder, 1994], Proposition 4.2.20. ////

6 Infinitesimal Robust Setup

For a very detailed introduction and motivation of robust statistics we refer to Chapter 1 of [Hampel et al., 1986]. A quick introduction to robustness is also given by [Huber, 1997]. In this section we introduce the infinitesimal robust setup which may be found in Subsection 4.2.1 of [Rieder, 1994]. For a more detailed introduction to this setup we also refer to [Bickel, 1981]. Let

$$\mathcal{U}(\theta) = \big\{U(\theta,r)\,\big|\, r\in[0,\infty)\big\} \tag{6.1}$$

be any system of neighborhoods $U(\theta, r)$ of radius $r \in [0, \infty)$ about P_θ such that

$$P_\theta \in U(\theta, r_1) \subset U(\theta, r_2) \subset \mathcal{M}_1(\mathcal{A}) \qquad 0 \le r_1 < r_2 < \infty \tag{6.2}$$

Within this work we restrict ourselves to (convex) contamination ($* = c$) and total variation ($* = v$) neighborhood systems $\mathcal{U}_*(\theta)$. [Rieder, 1994] also considers Hellinger ($* = h$), Kolmogorov ($* = \kappa$), Cramér-von Mises ($* = \mu$), Prokhorov ($* = \pi$) and Lévy ($* = \lambda$) neighborhood systems. In the cases $* = c, v$ the system $\mathcal{U}_*(\theta)$ consists of closed balls about P_θ that are defined for an arbitrary sample space,

$$U_*(\theta, r) = B_*(P_\theta, r) \qquad r \in [0, \infty) \tag{6.3}$$

where

$$B_c(P_\theta, r) = \big\{(1-r)_+ P_\theta + (1 \wedge r)\, Q \,\big|\, Q \in \mathcal{M}_1(\mathcal{A})\big\} \tag{6.4}$$

$$B_v(P_\theta, r) = \big\{Q \in \mathcal{M}_1(\mathcal{A}) \,\big|\, d_v(Q, P_\theta) \le r\big\} \tag{6.5}$$

with metric

$$d_v(Q, P_\theta) = \frac{1}{2}\int |dQ - dP_\theta| = \sup_{A \in \mathcal{A}} |Q(A) - P_\theta(A)| \tag{6.6}$$

and it holds $B_c(P_\theta, r) \subset B_v(P_\theta, r)$.

Remark 6.1 The observations $y_1, \dots, y_n$, which are i.i.d. under the null hypothesis P_θ, may now be allowed to follow any law $Q \in U_*(\theta, r)$, while still the parameter θ has to be estimated. Since the equation

$$Q = P_\theta + (Q - P_\theta) \tag{6.7}$$

involving the nuisance component $Q - P_\theta$, has multiple solutions θ, the parameter θ is obviously no longer identifiable. ////

Next we define p-dimensional tangents at P_θ and introduce simple perturbations of P_θ.

Definition 6.2 *For any dimension $p \in \mathbb{N}$ and exponent $\alpha = 2, \infty$, respectively, we define*

$$Z_\alpha^p(\theta) = \big\{\zeta \in L_\alpha^p(P_\theta) \mid \mathrm{E}_\theta\, \zeta = 0\big\} \tag{6.8}$$

The elements of $Z_2^p(\theta)$ respectively, $Z_\infty^p(\theta)$ are called square integrable *respectively,* bounded p-dimensional tangents at P_θ*. If a parametric model $\mathcal{P}$ is L_2 differentiable at θ, the L_2 derivative Λ_θ is called* parametric tangent.

Remark 6.3 p-dimensional tangents may also be defined for arbitrary exponents $\alpha \in [1, \infty]$; confer Definition 4.2.1 of [Rieder, 1994]. ////

Definition 6.4 *A sequence $Q_n(\zeta, .)$ of* simple perturbations *of P_θ along $\zeta \in Z_2^k(\theta)$ is given by*

$$dQ_n(\zeta, t) = \Big(1 + \frac{1}{\sqrt{n}} t^\tau \zeta_n\Big) dP_\theta \qquad |t| \le \frac{\sqrt{n}}{\sup_{P_\theta} |\zeta_n|} \tag{6.9}$$

where the approximating bounded tangents $\zeta_n \in Z^k_\infty(\theta)$ are chosen such that

$$\lim_{n\to\infty} \mathrm{E}_\theta\, |\zeta_n - \zeta|^2 = 0 \qquad \sup_{P_\theta} |\zeta_n| = o(\sqrt{n}\,) \tag{6.10}$$

Every $t \in \mathbb{R}^k$ is eventually admitted as a parameter value. In case of $\zeta \in Z^k_\infty(\theta)$ we may choose $\zeta_n = \zeta$; confer Remark 4.2.3 of [Rieder, 1994].

The contamination and total variation neighborhood systems cover simple perturbations along $Z^k_\infty(\theta)$ $(* = c)$ respectively, $Z^k_2(\theta)$ $(* = v)$, on the $1/\sqrt{n}$ scale. The following definition corresponds to Definition 4.2.6 of [Rieder, 1994].

Definition 6.5 *We call the neighborhood system $\mathcal{U}(\theta)$ about P_θ full if for every $\zeta \in Z^k_\infty(\theta)$ and $c \in (0,\infty)$ there exist some $r \in (0,\infty)$ and $n_0 \in \mathbb{N}$ such that*

$$t \in \mathbb{R}^k,\ |t| \le c,\ n \in \mathbb{N},\ n > n_0 \quad\Longrightarrow\quad Q_n(\zeta,t) \in U(\theta, r/\sqrt{n}\,) \tag{6.11}$$

In this context we also note Remark 4.2.7 of [Rieder, 1994] as part (a) of the following remark.

Remark 6.6 **(a)** With the $1/\sqrt{n}$ scaling, a neighborhood system is also called *infinitesimal*. For sample size $n \to \infty$, neighborhoods and simple perturbations are scaled down so, because, on the one hand, such deviations from the ideal model have nontrivial effects on statistical procedures, while, on the other hand, they cannot be detected surely by goodness-of-fit tests.

(b) The question, why infinitesimal contamination neighborhoods are shrinking at a rate of $\sqrt{n}$, is also answered in [Ruckdeschel, 2005] by constructing a Neyman-Pearson test for binomial probabilities to detect outliers. ////

Lemma 6.7 *The systems $\mathcal{U}_c(\theta)$ and $\mathcal{U}_v(\theta)$ are full and they cover simple perturbations along $Z^k_\infty(\theta)$ and $Z^k_2(\theta)$, respectively.*

PROOF: [Rieder, 1994], parts (c) and (v) of Lemma 4.2.8. ////

As the following lemma shows, the sequence of n-fold product measures $Q^n_n(\zeta,\,.\,)$ is LAN with asymptotically sufficient statistic $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \zeta(y_i)$ and asymptotic covariance $C = \mathrm{Cov}_\theta(\zeta)$.

Lemma 6.8 *For every convergent sequence $t_n \to t$ in $\mathbb{R}^k$, the simple perturbations $Q_n(\zeta,\,.\,)$ along $\zeta \in Z^k_2(\theta)$, as defined by (6.9) and (6.9), satisfy*

$$\lim_{n\to\infty} \sqrt{n}\,\Big\| \sqrt{dQ_n(\zeta,t_n)} - \sqrt{dP_\theta}\,\Big(1 + \tfrac{1}{2\sqrt{n}} t^\tau \zeta\Big)\Big\|_{\mathcal{L}^k_2} = 0 \tag{6.12}$$

and

$$\log \frac{dQ^n_n}{dP^n_\theta} = t^\tau \frac{1}{\sqrt{n}} \sum_{i=1}^n \zeta(y_i) - \tfrac{1}{2} t^\tau\, \mathrm{Cov}_\theta(\zeta) t + o_{P^n_\theta}(n^0) \tag{6.13}$$

PROOF: [Rieder, 1994], Lemma 4.2.4. ////

As a consequence of Lemma 6.8 together with Slutzky's lemma, the Cramér-Wold device and Le Cam's third lemma (Theorem 3.1) we get,

Proposition 6.9 *Consider an estimator S that is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$ and let $Q_n(\zeta, .)$ be a sequence of simple perturbations along $\zeta \in Z_2^k(\theta)$. Then*

$$\sqrt{n}\,(S_n - \theta)(Q_n^n(\zeta, t_n)) \underset{\mathrm{w}}{\longrightarrow} \mathcal{N}_k(\mathrm{E}_\theta\, \psi_\theta \zeta^\tau t,\, \mathrm{Cov}_\theta(\psi_\theta)) \tag{6.14}$$

for all convergent sequences $t_n \to t$ in $\mathbb{R}^k$.

Remark 6.10 Assume transforms $\tau \colon \mathbb{R}^k \to \mathbb{R}^p$ $(p \le k)$ which are differentiable at θ with bounded derivative $D = d\tau(\theta)$ of full rank p,

$$\tau(\theta + t) = \tau(\theta) + Dt + o(|t|) \qquad rkD = p \tag{6.15}$$

Then we get by the finite-dimensional delta method, setting $\eta_\theta = D\psi_\theta$,

$$\sqrt{n}\,\big(\tau \circ S_n - \tau(\theta)\big)(Q_n^n(\zeta, t_n)) \underset{\mathrm{w}}{\longrightarrow} \mathcal{N}_p(\mathrm{E}_\theta\, \eta_\theta \zeta^\tau t,\, \mathrm{Cov}_\theta(\eta_\theta)) \tag{6.16}$$

for all convergent sequences $t_n \to t$ in $\mathbb{R}^k$. ////

7 Optimal Influence Curves

7.1 Introduction

We now fix $\theta \in \Theta$ and define the following subclasses of one-dimensional bounded tangents

$$\mathcal{G}_c(\theta) = \big\{ q \in Z_\infty(\theta) \;\big|\; \inf\nolimits_{P_\theta} q \ge -1 \big\} \tag{7.1}$$

and

$$\mathcal{G}_v(\theta) = \big\{ q \in Z_\infty(\theta) \;\big|\; \mathrm{E}_\theta\, |q| \le 2 \big\} \tag{7.2}$$

By formally identifying $t^\tau \zeta = rq$, the simple perturbations along $\zeta \in Z_\infty^k(\theta)$ are, for $\sqrt{n} \ge -r \inf_{P_\theta} q$,

$$dQ_n(q, r) = dQ_n(\zeta, t) = \Big(1 + \frac{r}{\sqrt{n}} q\Big) dP_\theta \tag{7.3}$$

Lemma 7.1 *Given $q \in Z_\infty(\theta)$ and $r \in (0, \infty)$. Then, in the cases $* = c, v$, for every $n \in \mathbb{N}$ such that $\sqrt{n} \ge -r \inf_{P_\theta} q$,*

$$Q_n(q, r) \in B_*(P_\theta, r/\sqrt{n}\,) \quad \Longleftrightarrow \quad q \in \mathcal{G}_*(\theta) \tag{7.4}$$

PROOF: On identifying $t^\tau \zeta = rq$ this may be read off the parts (c) and (v) of the proof of Lemma 4.2.8 in [Rieder, 1994]. ////

In view of Proposition 6.9 and Remark 6.10 we obtain the following result.

Proposition 7.2 *Consider an estimator S that is asymptotically linear at P_θ with IC $\psi_\theta \in \Psi_2(\theta)$ and let $Q_n(\zeta,.)$ be a sequence of simple perturbations along $q \in Z_\infty(\theta)$. Moreover assume transforms $\tau\colon \mathbb{R}^k \to \mathbb{R}^p$ $(p \le k)$ which are differentiable at θ with bounded derivative $D = d\tau(\theta)$ of full rank p,*

$$\tau(\theta + t) = \tau(\theta) + Dt + o(|t|) \qquad rkD = p \tag{7.5}$$

and let $\eta_\theta = D\psi_\theta$ and

$$\rho_0 = \int \ell \, d\mathcal{N}_k(r \operatorname{E}_\theta \eta_\theta q, \operatorname{Cov}_\theta(\eta_\theta)) \tag{7.6}$$

(a) *If $\ell\colon \mathbb{R}^p \to [0,\infty]$ is lower semicontinuous then for all $r \in (0,\infty)$,*

$$\liminf_{n\to\infty} \int \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) \ge \rho_0 \tag{7.7}$$

(b) *If $\ell\colon \mathbb{R}^p \to [0,\infty]$ is continuous a.e. λ^p then for all $r \in (0,\infty)$,*

$$\lim_{M\to\infty} \lim_{n\to\infty} \int M \wedge \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) = \rho_0 \tag{7.8}$$

Proof: Consequence of Proposition 6.9 and Remark 6.10 together with

(a) the Lemma of Fatou in the version of Lemma A.2.1 of [Rieder, 1994].

(b) the continuous mapping theorem. ////

This leads us to the following limiting risk for ALEs, in the cases $* = c, v$,

$$\sup_{q \in \mathcal{G}_*(\theta)} \lim_{M\to\infty} \lim_{n\to\infty} \int M \wedge \ell\big(\sqrt{n}\,(\tau \circ S_n - \tau(\theta))\big)\, dQ_n^n(q,r) = \sup_{q \in \mathcal{G}_*(\theta)} \rho_0 \tag{7.9}$$

Choosing quadratic loss $\ell(z) = |z|^2$, we obtain the subsequent asymptotic mean square error (MSE) problems,

$$\operatorname{maxMSE}_\theta(\eta_\theta, r) := E_\theta|\eta_\theta|^2 + r^2 \omega_{*,\theta}(\eta_\theta)^2 = \min! \qquad \eta_\theta \in \Psi_2^D(\theta) \tag{7.10}$$

with

$$\omega_{*,\theta}(\eta_\theta) \;=\; \sup\big\{ |\operatorname{E}_\theta \eta_\theta q| \,\big|\, q \in \mathcal{G}_*(\theta) \big\} \tag{7.11}$$

where the radius $r \in (0,\infty)$ of the simple perturbations (7.3) is fixed. The solutions to this optimization problems are given in Subsection 7.4. The determination of the solutions is based on Langrange multiplier theorems derived in Appendix B of [Rieder, 1994] and canonically leads to the following Hampel type problem[1], with bound $b \in (0,\infty)$ fixed,

$$E_\theta|\eta_\theta|^2 = \min! \qquad \eta_\theta \in \Psi_2^D(\theta),\; \omega_{*,\theta}(\eta_\theta) \le b \tag{7.12}$$

Thus, the solutions to this Hampel type problems are given beforehand in Subsection 7.3. The standardized (infinitesimal) bias terms $\omega_{*,\theta}(\eta_\theta)$ that occur in the optimization problems are more or less explicitly calculated in Subsection 7.2.

[1] in allusion to the problem solved in Lemma 5 of [Hampel, 1968]

Remark 7.3 **(a)** Actually, we are interested in the following limiting risk

$$\lim_{M\to\infty}\lim_{n\to\infty}\sup_{Q\in U_*(\theta,r/\sqrt{n})}\int M\wedge\ell\big(\sqrt{n}\,(\tau\circ S_n-\tau(\theta))\big)\,dQ^n \tag{7.13}$$

Thus, it must be made sure, that at least for the optimal ICs, the interchanging of $\lim_M \lim_n$ and $\sup_q$ and the passage from the neighborhood submodel to full neighborhoods does not increase the asymptotic risk (7.9). Under additional assumptions on the optimal ICs, this goal can be achieved by suitable estimator contructions described in Chapter 6 of [Rieder, 1994].

(b) Since the normal distribution is fully specified by its first two moments, one might, analogously to pp. 197 of [Fraiman et al., 2001], think of the following general optimality problem

$$\sup_{q\in\mathcal{G}_*(\theta)} g\big(r\,\mathrm{E}_\theta\,\eta_\theta q,\mathrm{Cov}_\theta(\eta_\theta)\big)=\min! \qquad \eta_\theta\in\Psi_2^D(\theta) \tag{7.14}$$

for suitable functions g. By choosing $g(x_1,x_2)=|x_1|^2+\mathrm{tr}(x_2)$ and $g(x_1,x_2)=\infty\,\mathrm{I}_{\{|x_1|>b\}}(x_1)+\mathrm{tr}(x_2)$, respectively, this problem also covers the MSE and the Hampel type problem stated above. ////

To lighten the notation we drop the fixed parameter θ and write $\omega_*=\omega_{*,\theta}$ and $\eta=\eta_\theta$ as well as $\mathcal{G}_*=\mathcal{G}_*(\theta)$ and $\Psi_2^D=\Psi_2^D(\theta)$. Moreover let $\mathrm{E}=\mathrm{E}_\theta$ denote expectation, $\mathrm{Cov}=\mathrm{Cov}_\theta$ covariance, and $\inf_P$, $\sup_P$ the essential extrema, under $P=P_\theta$.

7.2 Bias Terms

The standardized bias terms ω_* for $*=c,v$ have the following general properties.

Lemma 7.4 *Let $*=c,v$ and $\eta\in L_1^p(P)$. Then*

$$\omega_*(\eta)=\omega_*(\eta-\mathrm{E}\,\eta) \tag{7.15}$$
$$\omega_*(\eta)=\sup\big\{\omega_*(e^\tau\eta)\,\big|\,e\in\mathbb{R}^p,\ |e|=1\big\} \tag{7.16}$$
$$\omega_c(\eta)\le\omega_v(\eta)\le 2\omega_c(\eta) \tag{7.17}$$

The terms ω_* are positively homogeneous, subadditive, hence convex on $L_1^p(P)$, and weakly lower semicontinuous on $L_2^p(P)$.

Proof: [Rieder, 1994], Lemma 5.3.2. ////

One gets the following explicit expressions for ω_*.

Proposition 7.5 *Let $\eta\in L_1(P)$ with $\mathrm{E}\,\eta=0$. Then*

$$\omega_c(\eta)=\ \sup\nolimits_P|\eta| \tag{7.18}$$
$$\omega_v(\eta)=\ \sup\big\{\sup\nolimits_P e^\tau\eta-\inf\nolimits_P e^\tau\eta\,\big|\,e\in\mathbb{R}^p,\ |e|=1\big\} \tag{7.19}$$

Proof: [Rieder, 1994], Proposition 5.3.3 (a). ////

Remark 7.6 For $\eta \in L_1(P)$ such that η is bounded and $\mathrm{E}\,\eta = 0$, it turns out that the standardized bias terms evaluated over full contamination and total variation balls do not exceed ω_* by more than the increase of some P essential to pointwise extrema; confer Lemma 5.3.4 of [Rieder, 1994]. ////

7.3 Minimum Trace Subject to Bias Bound

In this section we give the unique solutions to the Hampel type problems (7.12). For various aspects of this problem confer pp. 196 of [Rieder, 1994]. We first give the unique solution for $* = c$.

Theorem 7.7 **(a)** *In case $\omega_c^{\min} < b \le \omega_c(\eta_h)$, there exist some $a \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p\times k}$ such that the solution is of the form*

$$\tilde\eta = (A\Lambda - a)w \qquad w = \min\left\{1, \frac{b}{|A\Lambda - a|}\right\} \tag{7.20}$$

Conversely, if some $\tilde\eta \in \Psi_2^D$ is of form (7.20) for any $b \in (0,\infty)$, $a \in \mathbb{R}^p$, and $A \in \mathbb{R}^{p\times k}$, then $\tilde\eta$ is the solution, and the following representations hold,

$$a = Az \qquad 0 = \mathrm{E}(\Lambda - z)w \qquad D = A\,\mathrm{E}(\Lambda - z)(\Lambda - z)^\tau w \tag{7.21}$$

where $AD^\tau = DA^\tau \succ 0$.

(b) *It holds that*

$$\omega_c^{\min} = \min\left\{\frac{\mathrm{E}\,|A\Lambda - a|}{\operatorname{tr} AD^\tau}\;\middle|\; a \in \mathbb{R}^p,\ A \in \mathbb{R}^{p\times k}\setminus\{0\}\right\} \tag{7.22}$$

There exist $a \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times k}$ and $\bar\eta \in \Psi_2^D$ achieving $\omega_c^{\min} = b$, respectively. And then necessarily

$$\bar\eta = b\frac{A\Lambda - a}{|A\Lambda - a|} \qquad \text{on } \{A\Lambda \ne a\} \tag{7.23}$$

Moreover, $a = Az$ for some $z \in \mathbb{R}^k$, and $AD^\tau = DA^\tau \succeq 0$.

If $\bar\eta$ in addition is constant on $\{A\Lambda = a\}$, then it is the solution.

PROOF: [Rieder, 1994], Theorem 5.5.1. ////

Since the explicit bias terms for $* = v$ and $p \ge 2$ are difficult to handle, we state the unique solution only for $p = 1$.

Theorem 7.8 **(a)** *In case $\omega_v^{\min} < b \le \omega_v(\eta_h)$, there exist some $c \in (-b, 0)$ and $A \in \mathbb{R}^{1\times k}$ such that*

$$\tilde\eta = c \vee A\Lambda \wedge (c + b) \tag{7.24}$$

is the solution, and

$$\omega_v(\tilde\eta) = b \tag{7.25}$$

Conversely, if some $\tilde\eta \in \Psi_2^D$ is of form (7.24) for any $b \in (0,\infty)$, $c \in \mathbb{R}$, and $A \in \mathbb{R}^{1\times k}$, then $\tilde\eta$ is the solution, and the following representations hold,

$$\mathrm{E}(c - A\Lambda)_+ = \mathrm{E}(A\Lambda - (c+b))_+ \qquad D = \mathrm{E}\left[c \vee A\Lambda \wedge (c+b)\right]\Lambda^\tau \tag{7.26}$$

(b) *It holds that*

$$\omega_v^{\min} = \min\left\{ \frac{\mathrm{E}(A\Lambda)_+}{AD^\tau} \,\middle|\, A \in \mathbb{R}^{1\times k} \setminus \{0\} \right\} \tag{7.27}$$

There exist $A \in \mathbb{R}^{1\times k}$ and $\bar\eta \in \Psi_2^D$ achieving $\omega_v^{\min} = b$, respectively. And then necessarily

$$\bar\eta\,\mathrm{I}(A\Lambda \neq 0) = c\,\mathrm{I}(A\Lambda < 0) + (c+b)\,\mathrm{I}(A\Lambda > 0) \tag{7.28}$$

for some $c \in (-b, 0)$. In the case $k = 1$, the solution is

$$\bar\eta = b\,\mathrm{sign}(D)\left(\frac{P(\Lambda<0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda>0) - \frac{P(\Lambda>0)}{P(\Lambda\neq 0)}\,\mathrm{I}(\Lambda<0)\right) \tag{7.29}$$

Proof: [Rieder, 1994], Theorem 5.5.5. ////

7.4 Mean Square Error

In this section we give the solutions to the MSE problemss (7.10).

Theorem 7.9 **(a)** *The solutions to problem (7.10) for $* = c$ and $(* = v,\ p = 1)$, respectively, are unique.*

(b) *The solution to problem (7.10) and $* = c$ coincides with the solution of problem (7.12) and $* = c$, with $b \in (0,\infty)$ and $r \in (0,\infty)$ related by*

$$r^2 b = \mathrm{E}\left(|A\Lambda - a| - b\right)_+ \tag{7.30}$$

(c) *The solution to problem (7.10) and $(* = v,\ p = 1)$ coincides with the solution of problem (7.12) and $(* = v,\ p = 1)$, with $b \in (0,\infty)$ and $r \in (0,\infty)$ related by*

$$r^2 b = \mathrm{E}\left(c - A\Lambda\right)_+ \tag{7.31}$$

Proof: [Rieder, 1994], Theorem 5.5.7. ////

The following result may, in terms of statistical risk, be interpreted as an extension of the classical Cramér-Rao bound under quadratic loss, in which $\mathrm{tr}\,A = \mathrm{tr}\,\mathcal{I}^{-1}$.

Lemma 7.10 *It holds*

$$\mathrm{maxMSE}\,(\tilde\eta, r) = \mathrm{tr}\,AD^\tau \tag{7.32}$$

PROOF:
$* = c$: We define $Y := A(\Lambda - z)$. Then

$$\begin{aligned}\operatorname{maxMSE}(\tilde\eta, r) &= \mathrm{E}\,|\tilde\eta|^2 + r^2\omega_c^2(\tilde\eta)\\ &= \mathrm{E}\,|Y|^2 w^2 + r^2 b^2\\ &= \mathrm{E}\,Y^\tau\tilde\eta - \mathrm{E}\,Y^\tau Y w(1-w) + r^2b^2\\ &= \operatorname{tr} AD^\tau - \mathrm{E}\,Y^\tau Y w(1-w) + r^2b^2 \qquad \text{by (7.21)}\end{aligned}$$

Moreover

$$\begin{aligned}\mathrm{E}\,Y^\tau Y w(1-w) &= \mathrm{E}\,|Y|^2 \frac{b}{|Y|}\Big(1 - \frac{b}{|Y|}\Big)_+\\ &= b\,\mathrm{E}\,\big(|Y| - b\big)_+\\ &= r^2b^2 \qquad \text{by (7.30)}\end{aligned}$$

$* = v$, $p = 1$: We define $Y := A\Lambda \in \mathbb{R}$ and $w := 1 \wedge \max\{\frac{c}{Y}, \frac{c+b}{Y}\}$. Then $\tilde\eta = Yw$ and we get analogously to the case $* = c$

$$\operatorname{maxMSE}(\tilde\eta, r) = AD^\tau - \mathrm{E}\,Y^2 w(1-w) + r^2b^2 \qquad \text{by (7.26)} \tag{7.33}$$

Again

$$\begin{aligned}\mathrm{E}\,Y^2 w(1-w) &= \mathrm{E}\,Y^2 \max\Big\{\frac{c}{Y}, \frac{c+b}{Y}\Big\}\Big(1 - \max\Big\{\frac{c}{Y}, \frac{c+b}{Y}\Big\}\Big)_+ &\quad (7.34)\\ &= (c+b)\,\mathrm{E}\,\big(Y - (c+b)\big)_+ - c\,\mathrm{E}\,\big(c - Y\big)_+\\ &= r^2b^2 \qquad \text{by (7.26) and (7.31)}\end{aligned}$$

////

Remark 7.11 This correspondence for the asymptotic minimax MSE holds more generally and can be verified for the cases $* = c, v$, $t = 0, \varepsilon, \alpha$, $s = 0, e, 2$ considered in [Rieder, 1994]. Exceptions are the cases $* = h, t = 0, s = 0, e$ and $* = h, t = \alpha = 2, s = e$, where the optimal robust ICs are identical to η_h and $\operatorname{maxMSE}(\eta_h, r) = \operatorname{tr} D\mathcal{I}^{-1}D^\tau + r^2b^2$. ////

References

[Bickel, 1981] Bickel, P. (1981). Quelques aspects de la statistique robuste. In *Ecole d'ete de probabilites de Saint-Flour IX-1979, Lect. Notes Math. 876, 2-72* . 9, 10

[Bickel et al., 1998] Bickel, P. J., Klaassen, C. A., Ritov, Y., and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models.* Springer. 2

[Fraiman et al., 2001] Fraiman, R., Yohai, V. J., and Zamar, R. H. (2001). Optimal robust M-estimates of location. *Ann. Stat.*, 29(1):194–223. 15

[Hampel, 1968] Hampel, F. R. (1968). *Contributions to the theory of robust estimation.* Dissertation, University of California, Berkely, CA. 2, 14

[Hampel, 1974] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.*, 69:383–393. 9

[Hampel et al., 1986] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics. The approach based on influence functions.* Wiley Series in Probability and Mathematical Statistics. Wiley. 2, 7, 10

[Huber, 1964] Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Stat.*, 35:73–101. 9

[Huber, 1981] Huber, P. J. (1981). *Robust statistics.* Wiley Series in Probability and Mathematical Statistics. Wiley. 7

[Huber, 1997] Huber, P. J. (1997). *Robust statistical procedures. 2 ed.* CBMS-NSF Regional Conference Series in Applied Mathematics. 68. Philadelphia, PA: SIAM. ix, 67 p. 10

[Le Cam, 1960] Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. of California Publ.*, 3:37–98. 5

[Rieder, 1980] Rieder, H. (1980). Estimates derived from robust tests. *Ann. Stat.*, 8:106–115. 8, 9

[Rieder, 1994] Rieder, H. (1994). *Robust asymptotic statistics.* Springer Series in Statistics. Springer. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18

[Ruckdeschel, 2005] Ruckdeschel, P. (2005). A Motivation for $1/\sqrt{n}$-Shrinking-Neighborhoods? http://www.uni-bayreuth.de/departments/math/org/mathe7/RUCKDESCHEL/pubs/whysqn.pdf. 12

[van der Vaart, 1998] van der Vaart, A. (1998). *Asymptotic statistics.*, volume 3 of *Cambridge Series on Statistical and Probabilistic Mathematics.* Cambridge Univ. Press., Cambridge. 2

Author Index

Subject Index