Convergence of Robust Models

Dr. Matthias Kohl

Mathematics VII: Stochastics

UNIVERSITÄT BAYREUTH

Stochastiktage (GOCPS) 2008

Outline

UNIVERSITÄT BAYREUTH

Outline

UNIVERSITÄT BAYREUTH

Ideal Model

- parametric family of probability measures

$$\mathcal{P} = \{P_\theta \,|\, \theta \in \Theta\} \qquad \Theta \subset \mathbb{R}^k \text{ (open)}$$

  defined on some measurable space $(\Omega, \mathcal{A})$

- smoothly parameterized; i.e., L_2 differentiable at $\theta \in \Theta$ with L_2 derivative $\Lambda_\theta \in L_2^k(P_\theta)$, $\mathrm{E}_\theta \Lambda_\theta = 0$ and

- Fisher information of full rank

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

Influence Curves (ICs)

Definition

The set $\Psi_2(\theta)$ *of all* **square integrable ICs at P_θ** *consists of*

- *all* $\psi_\theta \in L_2^k(P_\theta)$ *which are*
- *centered:* $E_\theta\, \psi_\theta = 0$ *and*
- *Fisher consistent:* $E_\theta\, \psi_\theta \Lambda_\theta^\tau = \mathbb{I}_k$

(where $\mathbb{I}_k$ k-dimensional identity matrix)

AL Estimators

Definition

An asymptotic estimator $S_n\colon (\Omega^n, \mathcal{A}^n) \to (\mathbb{R}^k, \mathbb{B}^k)$ *is called* **asymptotically linear at P_θ** *if there is an*

- *IC* $\psi_\theta \in \Psi_2(\theta)$
- *with*

$$S_n = \theta + \frac{1}{n}\sum_{i=1}^{n} \psi_\theta(y_i) + o_{P_\theta^n}(n^{1/2})$$

Infinitesimal Neighborhoods

Convex contamination neighborhood (gross error model)

$$U_c(\theta, r_n) = \{(1 - r_n)_+ P_\theta + (1 \wedge r_n)\, Q \,\big|\, Q \in \mathcal{M}_1(\mathcal{A})\}$$

- $\mathcal{M}_1(\mathcal{A})$ set of all probability measures on $\mathcal{A}$
- radius $r_n := r/\sqrt{n}$ shrinks with sample size $n \in \mathbb{N}$ where $r \in [0, \infty)$

Unique asymptotic MSE Solution

Theorem 5.5.7 (b), Rieder (1994)

$$\tilde{\eta}_\theta = (A_\theta \Lambda_\theta - a_\theta) w \qquad w = \min\left\{1, \frac{b_\theta}{|A_\theta \Lambda_\theta - a_\theta|}\right\}$$

with Lagrange multipliers A_θ, a_θ and b_θ determined by

$$0 = E_\theta(\Lambda_\theta - z_\theta) w \qquad a_\theta = A_\theta z_\theta \tag{1}$$

$$\mathbb{I}_k = A_\theta \, E_\theta(\Lambda_\theta - z_\theta)(\Lambda_\theta - z_\theta)^\tau w \tag{2}$$

$$r^2 b_\theta = E_\theta \left(|A_\theta \Lambda_\theta - a_\theta| - b_\theta\right)_+ \tag{3}$$

Unique asymptotic MSE Solution

Theorem 5.5.7 (b), Rieder (1994)

$$\tilde{\eta}_\theta = (A_\theta \Lambda_\theta - a_\theta) w \qquad w = \min\left\{1, \frac{b_\theta}{|A_\theta \Lambda_\theta - a_\theta|}\right\}$$

with Lagrange multipliers A_θ, a_θ *and* b_θ *determined by*

$$0 = E_\theta (\Lambda_\theta - z_\theta) w \qquad a_\theta = A_\theta z_\theta \tag{1}$$

$$\mathbb{I}_k = A_\theta \, E_\theta (\Lambda_\theta - z_\theta)(\Lambda_\theta - z_\theta)^\tau w \tag{2}$$

$$r^2 b_\theta = E_\theta \left(|A_\theta \Lambda_\theta - a_\theta| - b_\theta \right)_+ \tag{3}$$

UNIVERSITÄT BAYREUTH

Extension of classical Cramér-Rao bound

Proposition 2.1.1, Kohl (2005)

$$\text{(as.)maxMSE}(\eta_\theta, r) \geq \text{(as.)maxMSE}(\tilde{\eta}_\theta, r) = \operatorname{tr} A_\theta$$

Classical Cramér-Rao bound:

$$\mathrm{Cov}(\eta_\theta) \succeq \mathrm{Cov}(\hat{\psi}_\theta) = \mathcal{I}_\theta^{-1} \qquad \text{where } \hat{\psi}_\theta = \mathcal{I}_\theta^{-1}\Lambda_\theta$$

Hence

$$\mathrm{MSE}\,(\eta_\theta) = \operatorname{tr}\mathrm{Cov}(\eta_\theta) \geq \operatorname{tr}\mathrm{Cov}(\hat{\psi}_\theta) = \operatorname{tr}\mathcal{I}_\theta^{-1} = \mathrm{MSE}\,(\hat{\psi}_\theta)$$

Estimator Construction

One-step construction of optimally robust estimator $\tilde{S}_n$

$$\tilde{S}_n = \hat{\theta} + \frac{1}{n} \sum_{i=1}^{n} \tilde{\eta}_{\hat{\theta}}(y_i)$$

where $\hat{\theta}$ is a uniformly consistent starting estimator.

Examples for initial estimators: Kolmogorov or Cramér von Mises minimum distance estimators; cf. Rieder (1994).

Estimator Construction

One-step construction of optimally robust estimator $\tilde{S}_n$

$$\tilde{S}_n = \hat{\theta} + \frac{1}{n} \sum_{i=1}^{n} \tilde{\eta}_{\hat{\theta}}(y_i)$$

where $\hat{\theta}$ is a uniformly consistent starting estimator.

Examples for initial estimators: Kolmogorov or Cramér von Mises minimum distance estimators; cf. Rieder (1994).

Outline

1. Asymptotic Theory of Robustness
   - Asymptotically Linear Estimators
   - Infinitesimal Neighborhoods
   - Optimally Robust Estimators
2. Convergence of Robust Models
   - Setup and Questions
   - Convergence of Robust Models

Setup

- Let

$$\mathcal{P}_\nu = \{P_{\nu,\theta} \,|\, \theta \in \Theta_\nu\} \subset \mathcal{M}_1(\mathcal{A}_\nu) \qquad \Theta_\nu \subset \mathbb{R}^k \text{ (open)}$$

  $(\nu \in \mathbb{N}_0)$ be a sequence of L_2-differentiable parametric families with L_2 derivatives $\Lambda_{\nu,\theta}$ and Fisher information of full rank $\mathcal{I}_{\nu,\theta}$.

- and consider

$$U_{\nu,c}(\theta, r_n) = \{(1 - r_n)_+ P_{\nu,\theta} + (1 \wedge r_n)\, Q_\nu \,\big|\, Q_\nu \in \mathcal{M}_1(\mathcal{A}_\nu)\}$$

UNIVERSITÄT BAYREUTH

Setup

- Let

$$\mathcal{P}_\nu = \{P_{\nu,\theta} \,|\, \theta \in \Theta_\nu\} \subset \mathcal{M}_1(\mathcal{A}_\nu) \qquad \Theta_\nu \subset \mathbb{R}^k \text{ (open)}$$

  ($\nu \in \mathbb{N}_0$) be a sequence of L_2-differentiable parametric families with L_2 derivatives $\Lambda_{\nu,\theta}$ and Fisher information of full rank $\mathcal{I}_{\nu,\theta}$.

- and consider

$$U_{\nu,c}(\theta, r_n) = \Big\{(1 - r_n)_+ P_{\nu,\theta} + (1 \wedge r_n)\, Q_\nu \,\Big|\, Q_\nu \in \mathcal{M}_1(\mathcal{A}_\nu)\Big\}$$

Questions

Questions:

Q1: $\mathcal{P}_\nu \approx \mathcal{P}_0$ for ν sufficiently large?

or even

Q2: $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$ for ν sufficiently large?

Q1: Convergence of Experiments; cf. Le Cam (1986)

Questions

Questions:

Q1: $\mathcal{P}_\nu \approx \mathcal{P}_0$ for ν sufficiently large?

or even

Q2: $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$ for ν sufficiently large?

Q1: Convergence of Experiments; cf. Le Cam (1986)

Questions

Questions:

Q1: $\mathcal{P}_\nu \approx \mathcal{P}_0$ for ν sufficiently large?

or even

Q2: $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$ for ν sufficiently large?

Q1: Convergence of Experiments; cf. Le Cam (1986)

Q2: Assumptions

Assume

$$\mathcal{L}_{P_{\nu,\theta_\nu}}(\gamma_\nu^{-1} G_\nu \Lambda_{\nu,\theta_\nu}) \xrightarrow{\mathrm{w}} \mathcal{L}_{P_{0,\theta_0}}(\Lambda_{0,\theta_0}) \qquad \text{as } \nu \to \infty \tag{4}$$

$$\gamma_\nu^{-2}\, tr\, \mathcal{I}_{\nu,\theta_\nu} \longrightarrow tr\, \mathcal{I}_{0,\theta_0} \qquad \text{as } \nu \to \infty \tag{5}$$

for $(\gamma_\nu)_{\nu\in\mathbb{N}} \subset (0,\infty)$ and orthogonal $(G_\nu)_{\nu\in\mathbb{N}} \subset \mathbb{R}^{k\times k}$.

Q2: Assumptions

Assume

$$\mathcal{L}_{P_{\nu,\theta_\nu}}(\gamma_\nu^{-1} G_\nu \Lambda_{\nu,\theta_\nu}) \xrightarrow{\mathrm{w}} \mathcal{L}_{P_{0,\theta_0}}(\Lambda_{0,\theta_0}) \qquad \text{as } \nu \to \infty \qquad (4)$$

$$\gamma_\nu^{-2}\, tr\, \mathcal{I}_{\nu,\theta_\nu} \longrightarrow tr\, \mathcal{I}_{0,\theta_0} \qquad \text{as } \nu \to \infty \qquad (5)$$

for $(\gamma_\nu)_{\nu\in\mathbb{N}} \subset (0,\infty)$ and orthogonal $(G_\nu)_{\nu\in\mathbb{N}} \subset \mathbb{R}^{k\times k}$.

Q2: Convergence of Robust Models

Theorem 2.4.1, Kohl (2005)

Assume (4) *and* (5) *and denote the Lagrange multipliers contained in the MSE optimal solution* $\tilde{\eta}_{\nu,\theta_\nu}$ *by* A_{ν,θ_ν}, a_{ν,θ_ν} *and* b_{ν,θ_ν}. *Then,*

$$\lim_{\nu\to\infty} \gamma_\nu^2 \, tr\, A_{\nu,\theta_\nu} = tr\, A_{0,\theta_0} \qquad \lim_{\nu\to\infty} \gamma_\nu b_{\nu,\theta_\nu} = b_{0,\theta_0}$$

In case A_{0,θ_0} *and* a_{0,θ_0} *are unique, then also*

$$\lim_{\nu\to\infty} \gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu = A_{0,\theta_0} \quad \text{and} \quad \lim_{\nu\to\infty} \gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} = a_{0,\theta_0}$$

Q2: Convergence of Robust Models

Theorem 2.4.1, Kohl (2005)

Assume (4) *and* (5) *and denote the Lagrange multipliers contained in the MSE optimal solution* $\tilde{\eta}_{\nu,\theta_\nu}$ *by* A_{ν,θ_ν}, a_{ν,θ_ν} *and* b_{ν,θ_ν}. *Then,*

$$\lim_{\nu\to\infty} \gamma_\nu^2 \, tr\, A_{\nu,\theta_\nu} = tr\, A_{0,\theta_0} \qquad \lim_{\nu\to\infty} \gamma_\nu b_{\nu,\theta_\nu} = b_{0,\theta_0}$$

In case A_{0,θ_0} *and* a_{0,θ_0} *are unique, then also*

$$\lim_{\nu\to\infty} \gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu = A_{0,\theta_0} \quad \text{and} \quad \lim_{\nu\to\infty} \gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} = a_{0,\theta_0}$$

Approximations

For ν sufficiently large, we have

$$
\begin{aligned}
\gamma_\nu^2 \, tr \, A_{\nu,\theta_\nu} \approx tr \, A_{0,\theta_0} \quad &\Longleftrightarrow \quad tr \, A_{\nu,\theta_\nu} \approx \gamma_\nu^{-2} \, tr \, A_{0,\theta_0} \\
\gamma_\nu b_{\nu,\theta_\nu} \approx b_{0,\theta_0} \quad &\Longleftrightarrow \quad b_{\nu,\theta_\nu} \approx \gamma_\nu^{-1} b_{0,\theta_0} \\
\gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu \approx A_{0,\theta_0} \quad &\Longleftrightarrow \quad A_{\nu,\theta_\nu} \approx \gamma_\nu^{-2} G_\nu A_{0,\theta_0} G_\nu^\tau \\
\gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} \approx a_{0,\theta_0} \quad &\Longleftrightarrow \quad a_{\nu,\theta_\nu} \approx \gamma_\nu^{-1} G_\nu a_{0,\theta_0}
\end{aligned}
$$

Note: To obtain an exact and not only an approximate IC we re-center and re-standardize the approximation.

UNIVERSITÄT BAYREUTH

Approximations

For ν sufficiently large, we have

$$\begin{aligned}
\gamma_\nu^2\, tr\, A_{\nu,\theta_\nu} \approx tr\, A_{0,\theta_0} \quad &\Longleftrightarrow \quad tr\, A_{\nu,\theta_\nu} \approx \gamma_\nu^{-2}\, tr\, A_{0,\theta_0}\\
\gamma_\nu b_{\nu,\theta_\nu} \approx b_{0,\theta_0} \quad &\Longleftrightarrow \quad b_{\nu,\theta_\nu} \approx \gamma_\nu^{-1} b_{0,\theta_0}\\
\gamma_\nu^2 G_\nu^\tau A_{\nu,\theta_\nu} G_\nu \approx A_{0,\theta_0} \quad &\Longleftrightarrow \quad A_{\nu,\theta_\nu} \approx \gamma_\nu^{-2} G_\nu A_{0,\theta_0} G_\nu^\tau\\
\gamma_\nu G_\nu^\tau a_{\nu,\theta_\nu} \approx a_{0,\theta_0} \quad &\Longleftrightarrow \quad a_{\nu,\theta_\nu} \approx \gamma_\nu^{-1} G_\nu a_{0,\theta_0}
\end{aligned}$$

Note: To obtain an exact and not only an approximate IC we re-center and re-standardize the approximation.

MSE-Inefficiency

Let χ be the re-centered and re-standardized approximating IC.

The MSE-Inefficiency of χ at radius $r \in [0, \infty)$ is defined as

$$relMSE_{\theta_0}(\chi, r) = \frac{E_{\theta_0} |\chi|^2 + r^2 [\sup_{P_{\theta_0}} |\chi|]^2}{tr\ A_{0,\theta_0}}$$

respectively as

$$relMSE_{\theta_\nu}(\chi, r) = \frac{E_{\theta_\nu} |\chi|^2 + r^2 [\sup_{P_{\theta_\nu}} |\chi|]^2}{tr\ A_{\nu,\theta_\nu}}$$

MSE-Inefficiency

Let χ be the re-centered and re-standardized approximating IC.

The MSE-Inefficiency of χ at radius $r \in [0, \infty)$ is defined as

$$relMSE_{\theta_0}(\chi, r) = \frac{E_{\theta_0} |\chi|^2 + r^2 [\sup_{P_{\theta_0}} |\chi|]^2}{tr\, A_{0,\theta_0}}$$

respectively as

$$relMSE_{\theta_\nu}(\chi, r) = \frac{E_{\theta_\nu} |\chi|^2 + r^2 [\sup_{P_{\theta_\nu}} |\chi|]^2}{tr\, A_{\nu,\theta_\nu}}$$

UNIVERSITÄT BAYREUTH

MSE-Inefficiency

Let χ be the re-centered and re-standardized approximating IC.

The MSE-Inefficiency of χ at radius $r \in [0, \infty)$ is defined as

$$relMSE_{\theta_0}(\chi, r) = \frac{E_{\theta_0} |\chi|^2 + r^2 [\sup_{P_{\theta_0}} |\chi|]^2}{tr\, A_{0,\theta_0}}$$

respectively as

$$relMSE_{\theta_\nu}(\chi, r) = \frac{E_{\theta_\nu} |\chi|^2 + r^2 [\sup_{P_{\theta_\nu}} |\chi|]^2}{tr\, A_{\nu,\theta_\nu}}$$

UNIVERSITÄT BAYREUTH

Example 1: Normal Approximation of Binomial

Normal approximation: (“$mp(1-p) \geq 9$”)

$$\mathit{Binom}\,(m, p) \approx \mathcal{N}\left(mp, mp(1-p)\right)$$

We have $\nu = m$ and

$$\gamma_m = \sqrt{\frac{m}{p(1-p)}} \qquad \text{and} \qquad G_m = 1$$

cf. Lemma 3.2.2 in Kohl (2005).

Example 1: Normal Approximation of Binomial

Normal approximation: ("$mp(1-p) \geq 9$")

$$Binom\,(m, p) \approx \mathcal{N}\left(mp, mp(1-p)\right)$$

We have $\nu = m$ and

$$\gamma_m = \sqrt{\frac{m}{p(1-p)}} \qquad \text{and} \qquad G_m = 1$$

cf. Lemma 3.2.2 in Kohl (2005).

Example 2: Poisson Approximation of Binomial

Poisson approximation: Let $\lambda = mp$ ("p small, m large"). Then,

$$Binom\,(m,p) \approx Pois\,(\lambda)$$

We have $\nu = m$ and

$$\gamma_m = \frac{m}{1-p} \qquad \text{and} \qquad G_m = 1$$

cf. Lemma 4.2.4 in Kohl (2005).

Example 2: Poisson Approximation of Binomial

Poisson approximation: Let $\lambda = mp$ (“p small, m large”). Then,

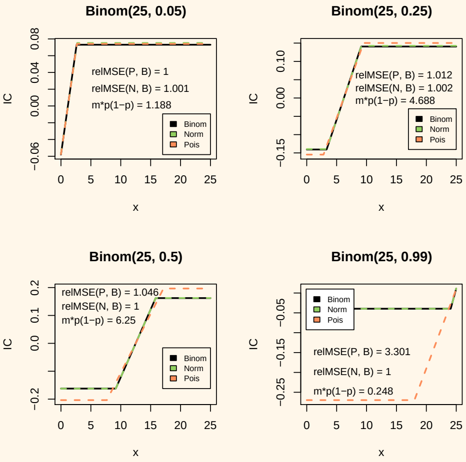
$$Binom\,(m, p) \approx Pois\,(\lambda)$$

We have $\nu = m$ and

$$\gamma_m = \frac{m}{1-p} \qquad \text{and} \qquad G_m = 1$$

cf. Lemma 4.2.4 in Kohl (2005).

Examples 1+2: Approximation of ICs for $r = 0.25$



Example 3: Exponential Scale and Gumbel Location

It holds

$$\mathcal{L}_{Gum(0,1)}\big(-\Lambda^{loc}_{(0,1)}\big) = \mathcal{L}_{Exp(1)}\big(\Lambda^{sc}_{1}\big), \qquad \mathcal{I}^{loc}_{Gum(0,1)} = \mathcal{I}^{sc}_{Exp(1)}$$

i.e., $\gamma_\nu \equiv 1$ and $G_\nu \equiv -1$; cf. Section 5.2.1 in Kohl (2005).

Hence,

$$relMSE\,(\chi, r) \equiv 1$$

Note: Further examples are given in Kohl (2005).

Example 3: Exponential Scale and Gumbel Location

It holds

$$\mathcal{L}_{Gum(0,1)}\left(-\Lambda^{loc}_{(0,1)}\right) = \mathcal{L}_{Exp(1)}\left(\Lambda^{sc}_{1}\right), \qquad \mathcal{I}^{loc}_{Gum(0,1)} = \mathcal{I}^{sc}_{Exp(1)}$$

i.e., $\gamma_\nu \equiv 1$ and $G_\nu \equiv -1$; cf. Section 5.2.1 in Kohl (2005).

Hence,

$$relMSE\,(\chi, r) \equiv 1$$

Note: Further examples are given in Kohl (2005).

Example 3: Exponential Scale and Gumbel Location

It holds

$$\mathcal{L}_{Gum(0,1)}\left(-\Lambda^{loc}_{(0,1)}\right) = \mathcal{L}_{Exp(1)}\left(\Lambda^{sc}_{1}\right), \qquad \mathcal{I}^{loc}_{Gum(0,1)} = \mathcal{I}^{sc}_{Exp(1)}$$

i.e., $\gamma_\nu \equiv 1$ and $G_\nu \equiv -1$; cf. Section 5.2.1 in Kohl (2005).

Hence,

$$relMSE\,(\chi, r) \equiv 1$$

Note: Further examples are given in Kohl (2005).

Summary

Convergence of Experiments

- $\mathcal{P}_\nu \approx \mathcal{P}_0$
- arbitrary loss functions
- arbitrary estimators
- computation of distance?

Convergence of Robust Models

- $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$
- L_2-risk
- optimally robust estimators
- MSE-inefficiency computable

Summary

Convergence of Experiments

- $\mathcal{P}_\nu \approx \mathcal{P}_0$
- arbitrary loss functions
- arbitrary estimators
- computation of distance?

Convergence of Robust Models

- $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$
- L_2-risk
- optimally robust estimators
- MSE-inefficiency computable

Summary

Convergence of Experiments

- $\mathcal{P}_\nu \approx \mathcal{P}_0$
- arbitrary loss functions
- arbitrary estimators
- computation of distance?

Convergence of Robust Models

- $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$
- L_2-risk
- optimally robust estimators
- MSE-inefficiency computable

Summary

Convergence of Experiments

- $\mathcal{P}_\nu \approx \mathcal{P}_0$
- arbitrary loss functions
- arbitrary estimators
- computation of distance?

Convergence of Robust Models

- $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$
- L_2-risk
- optimally robust estimators
- MSE-inefficiency computable

Summary

Convergence of Experiments

- $\mathcal{P}_\nu \approx \mathcal{P}_0$
- arbitrary loss functions
- arbitrary estimators
- computation of distance?

Convergence of Robust Models

- $U_{\nu,c}(\theta_\nu, r_n) \approx U_{0,c}(\theta_0, r_n)$
- L_2-risk
- optimally robust estimators
- MSE-inefficiency computable

UNIVERSITÄT BAYREUTH

Outlook

- Extension to convex risks; cf. Ruckdeschel and Rieder (2004)
- Find/Consider further (multivariate) examples
- Application: If optimal IC is hard to compute in one model try to find an approximating model where the optimal IC is easier to compute and use it as approximation.

Outlook

- Extension to convex risks; cf. Ruckdeschel and Rieder (2004)
- Find/Consider further (multivariate) examples
- Application: If optimal IC is hard to compute in one model try to find an approximating model where the optimal IC is easier to compute and use it as approximation.

Outlook

- Extension to convex risks; cf. Ruckdeschel and Rieder (2004)
- Find/Consider further (multivariate) examples
- Application: If optimal IC is hard to compute in one model try to find an approximating model where the optimal IC is easier to compute and use it as approximation.

Bibliography

M. Kohl.
Numerical Contributions to the Asymptotic Theory of Robustness. Dissertation, University of Bayreuth, 2005.

L. Le Cam
Asymptotic Methods in Statistical Decision Theory. Springer, 1986.

H. Rieder.
Robust Asymptotic Statistics. Springer, 1994.

P. Ruckdeschel and H. Rieder.
Optimal influence curves for general loss functions. Stat. Decis., 22: 201–223, 2004.