

Spread, estimators and nuisance parameters

EDWIN R. VAN DEN HEUVEL and CHRIS A.J. KLAASSEN*

Department of Mathematics and Institute for Business and Industrial Statistics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

A general spread inequality for arbitrary estimators of a one-dimensional parameter is given. This finite-sample inequality yields bounds on the distribution of estimators in the presence of finite- or infinite-dimensional nuisance parameters.

Keywords: finite-sample inequality; nuisance parameters; spread

Consider an arbitrary family \mathbf{P} of probability distributions P on a measurable space $(\mathbf{X}, \mathcal{A})$ and assume that this family is dominated by a σ -finite measure μ on $(\mathbf{X}, \mathcal{A})$. Let ν be a functional from \mathbf{P} to the real line \mathbb{R} , and let X be a random variable which takes values in \mathbf{X} and has its probability distribution P in \mathbf{P} . We are interested in estimation of $\nu(P)$ by an estimator $T = t(X)$ based on X .

We will consider the distribution of $c(P)(T - \nu(P))$ under P , where c is a functional from \mathbf{P} to $(0, \infty)$. This distribution may be anything and therefore very little can be said about it. On the other hand, it cannot be arbitrarily much concentrated for several possible $P \in \mathbf{P}$ simultaneously. To make this claim more precise, we will consider an average distribution function, where the average is taken over the family \mathbf{P} of probability distributions,

$$G(y) = \int_{\mathbf{P}} P(c(P)(T - \nu(P)) \leq y) d\tilde{W}(P), \quad y \in \mathbb{R}, \quad (1)$$

with the weight function \tilde{W} a probability measure on the measurable space $(\mathbf{P}, \mathcal{P})$. As stated in our Theorem 1 below, the distribution function G in (1) is at least as spread out as a certain distribution function K , notation $G \geq_1 K$. This means that any two quantiles of G are at least as far apart as the corresponding quantiles of K , i.e.

$$G^{-1}(u) - G^{-1}(v) \geq K^{-1}(u) - K^{-1}(v), \quad 0 < v \leq u < 1, \quad (2)$$

where the quantile function F^{-1} is defined by $F^{-1}(t) = \inf \{y: F(y) \geq t\}$. The partial ordering of one-dimensional distribution functions by spread has been introduced by Bickel and Lehmann (1979). The distribution function K depends only on \mathbf{P} , \tilde{W} , $\nu(\cdot)$ and $c(\cdot)$, but not on $t(\cdot)$. Therefore, K is a bound on the average distribution of an arbitrary estimator T according to the ordering of spread.

This so-called spread inequality (2) may be used to derive local asymptotic minimax

*To whom correspondence should be addressed. e-mail: vdheuvel@fwi.uva.nl

results; see Remark 1 below. Bootstrapping both the distribution of an estimator and the bound K , the performance of this estimator may be evaluated. For estimators of location this topic has been studied by Venetiaan (1994). The spread inequality also sharpens global Cramér–Rao inequalities; see Corollary 2.1 of Klaassen (1984) and formulae (2.4.20) and (2.4.29) of van den Heuvel (1996).

To prove the general spread inequality we will rewrite the distribution function G in (1). To that end, assume that the functional ν is $(\mathcal{P}, \mathcal{B})$ -measurable, with \mathcal{B} the Borel σ -field on \mathbb{R} , and c is $(\mathcal{P}, \mathcal{U})$ -measurable, with \mathcal{U} an arbitrary σ -field on $(0, \infty)$, and define the probability measure W on the measurable space $(\mathbb{R} \times (0, \infty), \mathcal{B} \times \mathcal{U})$ by $W(B \times U) = \tilde{W}(\{\nu(P), c(P) \in B \times U\})$, for every $B \in \mathcal{B}$, $U \in \mathcal{U}$. Then the distribution function G may be rewritten as

$$G(y) = \int_{\mathbb{R} \times (0, \infty)} \int_{\mathbf{P}_{\theta,z}} P(z(T - \theta) \leq y) d\tilde{W}(P|\nu(P), c(P)) = (\theta, z) dW(\theta, z), \tag{3}$$

where $\tilde{W}(\cdot|\nu(P), c(P)) = (\theta, z)$ is the conditional probability measure of \tilde{W} given $(\nu(P), c(P)) = (\theta, z)$ and $\mathbf{P}_{\theta,z}$ is the set of probability measures such that $(\nu(P), c(P)) = (\theta, z)$. If we assume that $\tilde{W}(\cdot|\nu(P), c(P))$ is a regular version of the conditional probability measure we may define the probability measure $P_{\theta,z}$ on $(\mathbf{X}, \mathcal{A})$ by

$$P_{\theta,z}(A) = \int_{\mathbf{P}_{\theta,z}} P(A) d\tilde{W}(P|\nu(P), c(P)) = (\theta, z), \quad A \in \mathcal{A}, \theta \in \nu(\mathbf{P}), z \in c(\mathbf{P}). \tag{4}$$

Note that $P_{\theta,z}$ is absolutely continuous with respect to μ , and denote its density by $p(\cdot|\theta, z)$. The distribution function G may be written as

$$G(y) = \int_{\mathbb{R} \times (0, \infty)} P_{\theta,z}(z(T - \theta) \leq y) dW(\theta, z). \tag{5}$$

If (ϑ, Z) is a random vector with probability distribution W , then G may be viewed as the distribution of $Z(t(X) - \vartheta)$, where the distribution of X given $(\vartheta, Z) = (\theta, z)$ is defined in (4). Consequently, the estimation problem is completely described by the joint distribution of (X, ϑ, Z) .

To obtain a bound on the distribution function G in (5) it suffices to show that G has a density g of the form

$$g(y) = \text{ES}1_{(y, \infty)}(Z(t(X) - \vartheta)), \quad y \in \mathbb{R}, \tag{6}$$

where S is a random variable based on the random vector (X, ϑ, Z) . In fact, S is the score statistic defined in (11) below. The bound K for distribution functions G satisfying (6) is defined by its inverse

$$K^{-1}(u) = \int_{\frac{1}{2}}^u \frac{1}{\int_s^1 H^{-1}(t) dt} ds, \quad 0 < u \leq 1, \tag{7}$$

with H the distribution function of the score statistic S . The validity of this lower bound is shown by rewriting (6) as

$$g(G^{-1}(s)) = \int_0^1 \psi(t)H^{-1}(t) dt, \quad 0 < s < 1, \tag{8}$$

with

$$\psi(t) = E(1_{(G^{-1}(s), \infty)}(Z(t(X) - \vartheta)) | S = H^{-1}(t)), \quad 0 < t < 1,$$

and by minimizing the right-hand side of (8) over all (critical) functions ψ , $0 \leq \psi \leq 1$, satisfying $1 - s = \int_0^1 \psi(t) dt$.

To obtain relation (6) a global L_1 -differentiability condition on the density of (X, ϑ, Z) suffices. The random vector (X, ϑ, Z) has density $f(x, \theta, z) \equiv p(x|\theta, z)w(\theta, z)$ on $\mathbf{X} \times \mathbb{R} \times (0, \infty)$ with respect to the measure $\nu \equiv \mu \times \text{Lebesgue} \times \lambda$ if W has a density w with respect to the σ -finite measure $\text{Lebesgue} \times \lambda$ on the measurable space $(\mathbb{R} \times (0, \infty), \mathcal{B} \times \mathcal{U})$. If there exists a function $h \in L_1(f)$ such that

$$\int_{\mathbf{X} \times \mathbb{R} \times (0, \infty)} |f(x, \theta + \epsilon z, z) - f(x, \theta, z) - \epsilon h(x, \theta, z)f(x, \theta, z)| d\nu(x, \theta, z) = o(\epsilon) \tag{9}$$

holds, then relation (6) is valid with $S = h(X, \vartheta, Z)$. In our Theorem 1 we will give sufficient conditions for relation (9) and the proof of this theorem shows how (9) implies (6) and hence our spread inequality.

Theorem 1. *Let \tilde{W} be a probability measure on the measurable space $(\mathbf{P}, \mathcal{P})$, ν a functional from \mathbf{P} to \mathbb{R} and c a functional from \mathbf{P} to $(0, \infty)$. Let $\tilde{W}(\cdot | (\nu(P), c(P)))$ be a regular version of the conditional probability measure given $(\nu(P), c(P))$, define the probability measure W on the measurable space $(\mathbb{R} \times (0, \infty), \mathcal{B} \times \mathcal{U})$ by $W(B \times U) = \tilde{W}((\nu(P), c(P)) \in B \times U)$, $B \in \mathcal{B}$, $U \in \mathcal{U}$, and assume that W has density w with respect to the σ -finite measure $\text{Lebesgue} \times \lambda$ on $(\mathbb{R} \times (0, \infty), \mathcal{B} \times \mathcal{U})$. Furthermore, let the probability measure $P_{\theta, z}$ be defined by (4) and assume that it has a density $p(\cdot | \theta, z)$ with respect to a σ -finite measure μ on $(\mathbf{X}, \mathcal{A})$. If the function $\theta \rightarrow f(x, \theta, z) \equiv p(x|\theta, z)w(\theta, z)$ is absolutely continuous with respect to Lebesgue measure on \mathbb{R} with Radon–Nikodym derivative $\dot{f}(x, \theta, z)$, for $\mu \times \lambda$ -almost all $(x, z) \in \mathbf{X} \times (0, \infty)$, if $\dot{f}(x, \theta, z)$ is $\mu \times \text{Lebesgue} \times \lambda$ -measurable and if*

$$\int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} z^{-1} |\dot{f}(x, \theta, z)| d\mu(x) d\theta d\lambda(z) < \infty \tag{10}$$

holds, then the distribution function G in (1) has density g satisfying (6) with S equal to

$$S = Z^{-1} \frac{\dot{f}}{f}(X, \vartheta, Z). \tag{11}$$

This implies that G is at least as spread out as K (see (2)) where K is defined in (7) and H is the distribution function of S in (11).

Proof. Since $\theta \rightarrow f(x, \theta, z)$ is absolutely continuous on \mathbb{R} for $\mu \times \lambda$ -almost all $(x, z) \in \mathbf{X} \times (0, \infty)$, we obtain

$$\limsup_{\epsilon \rightarrow 0} \int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} |\epsilon^{-1} \{f(x, \theta + \epsilon/z, z) - f(x, \theta, z)\}| d\mu(x) d\theta d\lambda(z)$$

$$\begin{aligned} &\leq \limsup_{\epsilon \rightarrow 0} \int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} \epsilon^{-1} \int_\theta^{\theta+\epsilon/z} |\dot{f}(x, t, z)| dt d\mu(x) d\theta d\lambda(z) \quad (12) \\ &= \int_0^\infty \int_{\mathbf{X}} \int_{\mathbb{R}} z^{-1} |\dot{f}(x, t, z)| dt d\mu(x) d\lambda(z) < \infty, \end{aligned}$$

in view of (10). From

$$1 - G(y + \delta) = \int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} 1_{(y, \infty)}(z(t(x) - \theta)) f(x, \theta - \delta/z, z) d\mu(x) d\theta d\lambda(z), \quad y \in \mathbb{R},$$

inequality (12) and Vitali’s theorem it follows that G is differentiable with derivative g satisfying

$$\begin{aligned} g(y) &= \lim_{\delta \rightarrow 0} \int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} 1_{(y, \infty)}(z(t(x) - \theta)) \delta^{-1} \{f(x, \theta, z) - f(x, \theta - \delta/z, z)\} d\mu(x) d\theta d\lambda(z) \\ &= \int_0^\infty \int_{\mathbb{R}} \int_{\mathbf{X}} 1_{(y, \infty)}(z(t(x) - \theta)) z^{-1} \dot{f}(x, \theta, z) d\mu(x) d\theta d\lambda(z). \end{aligned}$$

By Lemma 1.2.2 of Klaassen (1981), we obtain that $\{\theta \in \mathbb{R}: f(x, \theta, z) = 0, \dot{f}(x, \theta, z) \neq 0\}$ is a Lebesgue null set, for $\mu \times \lambda$ -almost all $(x, z) \in \mathbf{X} \times (0, \infty)$. This yields relation (6) with S given in (11). Now following the proof of Theorem 1.1 of Klaassen (1989a) from here on we obtain our result, via an argument as in the Neyman–Pearson lemma, as indicated below formula (7) above. □

Theorem 1 is a generalization of Theorem 1.1 of Klaassen (1989a). The requirement of the latter that $p(x|\theta, z)$ and $w(\theta, z)$ should be both absolutely continuous in θ is replaced here by the slightly weaker condition that $p(x|\theta, z)w(\theta, z)$ is absolutely continuous in θ . Taking \mathbf{P} to be one-dimensional parametric with parameter θ and taking the random variable $c(P) = Z$ to be degenerate at the constant a we obtain Theorem 1.1 of Klaassen (1989a). Our formulation is also more general in the sense that models with nuisance parameters are incorporated. In fact, both parametric models $\mathbf{P} = \{P_{\theta, \eta}: \theta \in \Theta, \eta \in \mathbf{H}\}$, $\Theta \subset \mathbb{R}$, $\mathbf{H} \subset \mathbb{R}^k$, and semiparametric models $\mathbf{P} = \{P_{\theta, F}: \theta \in \Theta, F \in \mathbf{F}\}$, $\Theta \subset \mathbb{R}$, for \mathbf{F} a set of distribution functions F , are included. This means that a spread inequality for arbitrary estimators of θ in the presence of finite- or infinite-dimensional (unknown) nuisance parameters is contained in Theorem 1. We will illustrate this via a parametric model $\mathbf{P} = \{P_{\theta, \eta}: \theta \in \Theta, \eta \in \mathbf{H}\}$ in the following example.

Example 1. Let X_1, X_2, \dots, X_n be independent identically normally distributed random variables with mean θ and standard deviation η . Our model \mathbf{P} is parametric with parameter space $\mathbb{R} \times (0, \infty)$, hence weight functions \tilde{W} on \mathbf{P} can be defined via distributions W on $\mathbb{R} \times (0, \infty)$. Here we choose $\nu(P) = \theta$ and $c(P) = \sqrt{n}/\eta = z$. The problem is estimating the parameter θ by an estimator $T = t(X_1, X_2, \dots, X_n)$ in the presence of the nuisance parameter η . This problem is determined by (X, ϑ, Z) , where the conditional distribution of $X = (X_1, X_2, \dots, X_n)^T$ given $(\vartheta, Z) = (\theta, z)$ has density

$$p(x|\theta, z) = \prod_{i=1}^n \frac{z}{\sqrt{n}} \phi\left(\frac{z}{\sqrt{n}}(x_i - \theta)\right), \quad x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n,$$

with ϕ the standard normal density. Let the distribution $W_n = W$ of (ϑ, Z) be such that the corresponding weight function in (θ, η) is independent of the sample size n , with the conditional distribution of ϑ given η normal with mean zero and variance $\eta^2\sigma^2$. More specifically, let W_n have density $w_n(\theta, z) = \sigma^{-1}z^{-1}\phi(\sigma^{-1}n^{-1/2}z\theta)w(n^{1/2}z^{-1})$ with respect to Lebesgue measure on $\mathbb{R} \times (0, \infty)$ with w a fixed density on $(0, \infty)$. Now, the score statistic S in (11) becomes

$$S = \frac{Z}{n} \sum_{i=1}^n (X_i - \vartheta) - \frac{Z}{\sigma^2 n} \vartheta.$$

The conditional distribution of S given $Z = z$ is equal to a normal distribution with mean zero and variance $1 + \sigma^{-2}n^{-1}$, and hence $H(y) = \Phi((1 + \sigma^{-2}n^{-1})^{-1/2}y)$. Furthermore, $\int_s^1 H^{-1}(t) dt = (1 + \sigma^{-2}n^{-1})^{1/2}\phi(\Phi^{-1}(s))$ and therefore $K(y) = \Phi((1 + \sigma^{-2}n^{-1})^{1/2}y)$. Consequently, Theorem 1 yields

$$G(\cdot) \geqslant_1 \Phi((1 + \sigma^{-2}n^{-1})^{1/2}\cdot), \tag{13}$$

where

$$G(y) = \int_0^\infty \int_{\mathbb{R}} P_{\theta,z}(z(t(X_1, X_2, \dots, X_n) - \theta) \leqslant y) \frac{1}{\sigma z} \phi\left(\frac{z}{\sigma\sqrt{n}}\theta\right) w_2\left(\frac{\sqrt{n}}{z}\right) d\theta dz. \tag{14}$$

Note that $T_n = t(X_1, X_2, \dots, X_n) = (n + \sigma^{-2})^{-1} \sum_{i=1}^n X_i$ is optimal, in the sense that equality is attained in (13). Consequently, our spread inequality (2) (see also Theorem 1) is sharp in this case. However, the conditional distribution of $Z(T_n - \vartheta)$ given $(\vartheta, Z) = (\theta, z)$ is normal with mean $-z(1 + \sigma^2n)^{-1}\theta$ and variance $(1 + \sigma^{-2}n^{-1})^{-2}$. This means that the optimal estimator T_n is biased, but asymptotically consistent. The biasedness is caused by the weight function or prior $w_n(\theta, z) = \sigma^{-1}z^{-1}\phi(\sigma^{-1}n^{-1/2}z\theta)w(n^{1/2}z^{-1})$, which is not uninformative. If we let σ tend to infinity then the prior tends to the uninformative prior, i.e. uninformative with respect to θ . Furthermore, $T_n = (n + \sigma^{-2})^{-1} \sum_{i=1}^n X_i$ converges to the sample mean $n^{-1} \sum_{i=1}^n X_i$, as $\sigma \rightarrow \infty$. Indeed, the sample mean normed by Z , i.e. $Z(n^{-1} \sum_{i=1}^n X_i - \vartheta)$, is standard normal as is the bound K in the limit.

Remark 1. Local asymptotic lower bounds may be obtained from Theorem 1 by choosing the weight function appropriately and by subsequently taking limits for sample size tending to infinity. For parametric models of n observations $\mathbf{P}_n = \{P_{\theta,\eta}^{(n)}; \theta \in \Theta, \eta \in \mathbf{H}\}$, $\Theta \subset \mathbb{R}$, $\mathbf{H} \subset \mathbb{R}$, we first identify weight functions \tilde{W}_n on \mathbf{P}_n via distributions W_n on $\Theta \times \mathbf{H}$, as in Example 1. Let W_n have density

$$a_n^2 w(a_n(\theta - \theta_0), a_n(\eta - \eta_0)), \tag{15}$$

where w is a density on \mathbb{R}^2 , $a_n = c(P_{\theta,\eta}^{(n)})$ is independent of θ and η , and $a_n \rightarrow \infty$ as $n \rightarrow \infty$. Under the same kind of conditions as in Theorem 4.1 of Klaassen (1989b) this yields a local asymptotic spread inequality at the parameter point (θ_0, η_0) . The corresponding asymptotic

bound is of the form (7), where H is the convolution of the limit distribution function of the efficient score function with respect to the parameter of interest θ and the nuisance parameter η by another distribution function. Choosing w appropriately and taking suitable limits we obtain a bound K which in local asymptotic normality situations is the same as in the convolution theorem and the local asymptotic minimax theorem.

For semiparametric models $\mathbf{P}_n = \{P_{\theta, \eta}^{(n)}: \theta \in \Theta, \eta \in \mathbf{H}\}$, $\Theta \subset \mathbb{R}$, \mathbf{H} infinite-dimensional, an asymptotic spread inequality at (θ_0, η_0) may be obtained by deriving spread inequalities for two-dimensional submodels through (θ_0, η_0) and subsequently maximizing these bounds over all possible submodels. For more details and examples of local asymptotic spread inequalities for parametric and semiparametric models see Sections 2.5, 2.6 and 4.2 of van den Heuvel (1996), and for a review of the theory of efficient estimation in semiparametric models based on the convolution theorem see Bickel *et al.* (1993).

Acknowledgements

We would like to thank the referees for their helpful comments. The research of the first author was financed by the Netherlands Organization for the Advancement of Scientific Research (NWO).

References

- Bickel, P.J. and Lehmann, E.L. (1979) Descriptive statistics for nonparametric models. IV. Spread. In J. Jurečková (ed.), *Contributions to Statistics. Jaroslav Hájek Memorial Volume*, pp. 33–40. Prague: Academia.
- Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: Johns Hopkins University Press.
- Klaassen, C.A.J. (1981) *Statistical Performance of Location Estimators*, Mathematical Centre Tract 133. Amsterdam: Mathematisch Centrum.
- Klaassen, C.A.J. (1984) Location estimators and spread. *Ann. Statist.*, **12**, 311–321.
- Klaassen, C.A.J. (1989a) Estimators and spread. *Ann. Statist.*, **17**, 859–867.
- Klaassen, C.A.J. (1989b) The asymptotic spread of estimators. *J. Statist. Plann. Inference*, **23**, 267–285.
- van den Heuvel, E.R. (1996) Bounds for statistical estimation in semiparametric models. Ph.D. thesis, University of Amsterdam.
- Venetiaan, S.A. (1994) Bootstrap bounds. Ph.D. thesis, University of Amsterdam.

Received November 1994 and revised September 1996.