

Second-order asymptotics for the number of times an estimator is more than ε from its target value

Nils Lid Hjort*, Grete Fenstad

Department of Statistics, 1 South Parks Road, Oxford OX1 3TG, UK

Department of Mathematics, P.B. 1053 Blindern, University of Oslo, N-0316 Oslo, Norway

Received 21 January 1993; revised 23 September 1994

Abstract

Suppose $\{\hat{\theta}_n: n \geq 1\}$ is a strongly consistent sequence of estimators for a parameter θ , where $\hat{\theta}_n$ is based on the first n observations. Consider Q_ε , the number of times $|\hat{\theta}_n - \theta| \geq \varepsilon$. In another paper (Hjort and Fenstad, *Ann. Statist.* **20** (1992) 469–489) we have shown that $\varepsilon^2 Q_\varepsilon$ has a limit distribution as $\varepsilon \rightarrow 0$, depending only on σ , the standard deviation of the limit distribution for $\sqrt{n}(\hat{\theta}_n - \theta)$, under natural regularity conditions. The present paper investigates some second-order asymptotics for differences between Q_ε -variables. The limit of $E(Q_{1,\varepsilon} - Q_{2,\varepsilon})$ is calculated in cases where $EQ_{1,\varepsilon}/EQ_{2,\varepsilon}$ goes to 1, leading to a notion of ‘asymptotic relative deficiency’ in cases where the asymptotic relative efficiency is 1. This is used to distinguish between competing estimators with identical limit distributions. Thus using denominator $n - \frac{1}{3}$ in the familiar formula for estimating a normal variance is better than both n and $n - 1$ and indeed all other choices, for example, in the sense of leading to the smallest possible expected number of ε -errors. Results of this type are found in a selection of familiar estimation problems, using limit results for expected differences, and are compared to corresponding asymptotic relative deficiency analysis in the sense of Hodges and Lehmann. Some second-order distributional results are reached as well. It is shown how ε times a Q_ε -difference tends to a variable which is related to some exponential distributions associated with Brownian motion, and that have recently been investigated by Hjort and Khasminskii (*Stochastic Process. Appl.* **47** (1993) 229–247).

Keywords: Asymptotic relative deficiency; Asymptotic relative efficiency; Comparison of estimators; Loss function; Second-order optimality; Total relative time for Brownian motion; The number of ε -misses

1. Introduction

Suppose $X_1, X_2, \dots$ is a sequence of independent observations from some distribution F , and let $\theta = \theta(F)$ be some parameter of interest. Assume that $\hat{\theta}_n$ is based on the

* Corresponding author. Tel.: 47 22 85 59 54. Fax: 47 22 85 43 49. E-mail: nils@math.uio.no.

first n data points, and that this estimator sequence is strongly consistent, i.e. $\hat{\theta}_n \rightarrow \theta$ almost surely. Then Q_ε , the number of cases where $|\hat{\theta}_n - \theta| \geq \varepsilon$, is finite almost surely. Under natural regularity conditions, which include existence of a normal $(0, \sigma^2)$ limit for $\sqrt{n}(\hat{\theta}_n - \theta)$, Hjort and Fenstad (1992, Section 7) have shown that

$$\varepsilon^2 Q_\varepsilon \rightarrow_d Q = \text{Leb}\{s \geq 0: |W(s)| \geq s/\sigma\} = \int_0^\infty I\{|W(s)| \geq s/\sigma\} ds, \quad (1.1)$$

in which Leb is Lebesgue measure on the half-line, $W(\cdot)$ is a Brownian motion process, and $I\{\dots\}$ denotes an indicator function. They also show that $\varepsilon^2 EQ_\varepsilon \rightarrow EQ$, which simply is equal to σ^2 . Suppose $\sqrt{n}(\hat{\theta}_{n,j} - \theta) \rightarrow_d N\{0, \sigma_j^2\}$ for two competing estimator sequences, and let $Q_{j,\varepsilon}$ be their accompanying number of ε -misses variables. Then a natural asymptotic measure of relative efficiency is

$$\text{a.r.e.} = \lim_{\varepsilon \rightarrow 0} \frac{EQ_{1,\varepsilon}}{EQ_{2,\varepsilon}} = \frac{\sigma_1^2}{\sigma_2^2}. \quad (1.2)$$

This is also the traditional formula for a.r.e. based on calculations of asymptotic ratios n_1/n_2 in which n_1 and n_2 are the sample sizes needed by methods 1 and 2 to achieve the same level of mean squared error precision. Hjort and Fenstad (1992) also give generalisations of (1.1) and (1.2) to multi-dimensional parameters and general distance measures, and reach limit results for Q_ε -variables in more involved problems like nonparametric estimation of distribution functions and densities.

The Q_ε -based criterion (1.2) is only a ‘first-order measure’ and cannot distinguish between estimator sequences with identical limit distributions, in which case typically $\varepsilon^2(Q_{1,\varepsilon} - Q_{2,\varepsilon}) \rightarrow 0$ and $Q_{1,\varepsilon}/Q_{2,\varepsilon} \rightarrow 1$ in probability. Our aim in this paper is to develop some ‘second-order’ theory for *differences* of Q_ε ’s, which then can be applied to single out the ‘best’ estimator in a class of estimators with the same limit distributions. Being ‘best’ here means having the smallest possible expected number of ε -errors, in the limit as ε goes to zero. In particular we intend to solve comparison problems for unbiased minimum variance (UMV) estimators versus maximum likelihood (ML) estimators versus Bayes estimators in some familiar cases where they disagree.

This is in a spirit similar to Hodges and Lehmann (1970) who worked with a measure of asymptotic relative deficiency, expressed as the limit of sample size difference $n_1 - n_2$ in cases where $n_1/n_2 \rightarrow 1$; see also Lehmann (1983, Ch. 5.2). In Section 2 formulae are found for a Q_ε -based

$$\text{a.r.d.} = \lim_{\varepsilon \rightarrow 0} E(Q_{1,\varepsilon} - Q_{2,\varepsilon}), \quad (1.3)$$

in the simple case of estimating a mean parameter, using estimators of the Bayesian variety $(n/(n+c))\bar{X}_n + (c/(n+c))d$. In this a.r.d. formula the underlying skewness of the X_i ’s enters in a natural way, whereas it does not in Hodges and Lehmann’s calculations. The a.r.d. formula makes it possible to show that $n - \frac{1}{3}$ is the superior

choice of denominator in the familiar $\sum_{i=1}^n (X_i - \bar{X}_n)^2 / (n + c)$ formula for a normal variance, for example. This and other mean-parameter related examples are treated in Section 3. A couple of technically more demanding problems are included in Section 4. We are able to show that $(\bar{X}_n)^2 + \sigma_n^2/n$, as an estimator for the squared normal mean, can be expected to exhibit fewer ε -errors than all other estimators of the form $(\bar{X}_n)^2 + d\sigma_n^2/n$, for example.

Comparing estimation methods in terms of EQ_ε can be phrased in decision theoretic terms, where the underlying loss function is equal to the number of ε -errors for the full sequence of estimates. This point is briefly discussed in Section 5, along with some Bayesian considerations.

The theory and applications presented in Sections 2–5 are only concerned with *expected values* of Q_ε -variables. This is reflected in the tools used, namely Edgeworth expansions and further Taylor-type approximations to probabilities. Section 6 presents a brief addendum and is concerned with second-order *distributional* aspects of $Q_{1,\varepsilon} - Q_{2,\varepsilon}$ differences. In the structurally simplest case of estimating a mean parameter it turns out that

$$\varepsilon^2(Q_{1,\varepsilon} - Q_{2,\varepsilon}) \rightarrow_p 0 \text{ and } Q_{1,\varepsilon}/Q_{2,\varepsilon} \rightarrow_p 1 \text{ while } \varepsilon(Q_{1,\varepsilon} - Q_{2,\varepsilon}) \rightarrow_d A - B,$$

say, where A and B are certain total relative time variables associated with Brownian motion, with distributions that are exponential or mixtures of exponentials and unit point masses at zero; see Hjort and Khasminskii (1993) for a separate account on these.

Our paper is mainly reporting on an investigation into one particular way of distinguishing between first-order equivalent sequences. We do not claim that the a.r.d. criterion (1.3) on which our second-order comparisons are based is statistically more natural than other criteria, but it is interesting that it can be computed at all, and to see how it fares compared to other second-order criteria, of which several others exist. In addition to the a.r.d. calculations of Hodges and Lehmann and ours of the present paper, Rao (1962) and others have worked with asymptotic sufficiency, Pfanzagl (1973) and others with coverage probabilities and median unbiasedness, and Ghosh and Subramanyam (1974) and others with expansions for mean squared errors. A good source for further information is the discussion of the paper by Berkson (1980). Let us finally point out that techniques of the present paper can be used to establish expansion results for coverage probabilities of the type $p_n = \Pr\{|\hat{\theta}_n - \theta| \leq k/\sqrt{n}\}$. These results, which in a way are simpler than those of Sections 2–5, could then be used to compare different estimators with the same limit distribution. Such comparisons would however depend on the value of k .

2. General results for estimating a mean

Suppose $X_1, X_2, X_3, \dots$ are i.i.d. with $EX_i = \xi$, $\text{Var } X_i = \sigma^2$, and skewness $E(X_i - \xi)^3/\sigma^3 = \gamma$. The most useful and directly informative results are those of

Propositions 1 and 2 below, but for technical reasons it is convenient to start with the following lemma.

Lemma. Let a be positive, and consider $Q_\varepsilon(c)$, the number of times, among $n \geq a/\varepsilon^2$, where $|(n/(n+c))\bar{X}_n - \xi| \geq \varepsilon$. Suppose that X_i has finite fourth-order moment and that its distribution is nonlattice, and write $\phi(\cdot)$ for the standard normal density. Then

$$\begin{aligned} \lambda_a(c) &= \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon(c) - Q_\varepsilon(0)\} \\ &= 2 \left[\frac{\xi^2}{\sigma^2} c^2 - \left(2 - \frac{2\gamma\xi}{3\sigma} \right) c \right] \int_{\sqrt{a}/\sigma}^{\infty} \phi(y) dy - \frac{2\gamma\xi}{3} \frac{\sqrt{a}}{\sigma} \phi(\sqrt{a}/\sigma) c. \end{aligned} \quad (2.1)$$

Proof. We choose to work with $Q_\varepsilon(c)$ in the form

$$Q_\varepsilon(c) = \sum_{n \geq am} I \left\{ \left| \frac{n}{n+c} \bar{X}_n - \xi \right| \geq \frac{1}{\sqrt{m}} \right\}, \quad (2.2)$$

writing $m = 1/\varepsilon^2$. Note that $Q_\varepsilon(c)$ is finite a.s. by the strong law of large numbers. Consider $T_n = \sqrt{(n/(n+c))}(\bar{X}_n - \xi)/\sigma$. Inserting $\bar{X}_n = \xi + \sigma T_n/\sqrt{n}$ the indicator function term in (2.2) can be written $I\{T_n \leq l(c) \text{ or } T_n \geq r(c)\}$, where

$$l(c) = -\frac{1}{\sigma} \sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}} \frac{c\xi}{\sigma} - \frac{1}{\sqrt{nm}} \frac{c}{\sigma}, \quad r(c) = \frac{1}{\sigma} \sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}} \frac{c\xi}{\sigma} + \frac{1}{\sqrt{nm}} \frac{c}{\sigma}. \quad (2.3)$$

Letting $s = n/m$ we have $l(0) = -u$, $l(c) = -u + b_1/\sqrt{n} - b_2/n$, $r(0) = u$, $r(c) = u + b_1/\sqrt{n} + b_2/n$, in which $u = \sqrt{s}/\sigma$, $b_1 = c\xi/\sigma$, and $b_2 = c\sqrt{s}/\sigma$. Since

$$\begin{aligned} Q_\varepsilon(c) - Q_\varepsilon(0) &= \sum_{n \geq am} [I\{T_n \leq l(c) \text{ or } T_n \geq r(c)\} - I\{T_n \leq l(0) \text{ or } T_n \geq r(0)\}], \end{aligned} \quad (2.4)$$

its expected value can be written

$$E\{Q_\varepsilon(c) - Q_\varepsilon(0)\} = \sum_{n \geq am} [\{G_n(l(c)) - G_n(l(0))\} - \{G_n(r(c)) - G_n(r(0))\}], \quad (2.5)$$

in terms of the cumulative distribution function G_n of T_n .

Under the nonlattice assumption there is a Cramér–Edgeworth expansion for G_n of the form

$$G_n(t) = \Phi(t) - \frac{1}{6} \frac{\gamma}{\sqrt{n}} A(t) \phi(t) + \frac{1}{n} R(t) \phi(t) + O(n^{-3/2}), \quad (2.6)$$

where $A(t) = t^2 - 1$ and $R(t)$ is a certain polynomial of degree five; see for example Barndorff-Nielsen and Cox (1989, Ch. 4). In view of (2.5) we need to approximate terms of the type $G_n(t + \delta) - G_n(t)$ where δ is of the order of $1/\sqrt{n}$. Using Taylor

expansions one finds that

$$G_n(t + \delta) - G_n(t) = \phi(t)\delta - \frac{1}{2}t\phi(t)\delta^2 + \frac{1}{6}\frac{\gamma}{\sqrt{n}}B(t)\phi(t)\delta + O(n^{-3/2}),$$

in which $B(t)\phi(t)$ is the derivative of $-A(t)\Phi(t)$, i.e. $B(t) = t^3 - 3t$, and the O -term is uniform in t . This leads to

$$G_n(-u + b_1/\sqrt{n} - b_2/n) - G_n(-u) = \phi(u)b_1/\sqrt{n} + \phi(u) \times \left\{ -b_2 + \frac{1}{2}ub_1^2 - \frac{1}{6}\gamma b_1 B(u) \right\}/n,$$

$$G_n(u + b_1/\sqrt{n} + b_2/n) - G_n(u) = \phi(u)b_1/\sqrt{n} + \phi(u) \left\{ b_2 - \frac{1}{2}ub_1^2 + \frac{1}{6}\gamma b_1 B(u) \right\}/n.$$

The summand in (2.5) can accordingly be expressed, apart from $O(n^{-3/2})$ terms, as

$$\begin{aligned} \frac{1}{n}\phi(u) \left\{ -2b_2 + ub_1^2 - \frac{1}{3}\gamma b_1 B(u) \right\} &= \frac{1}{m}\phi\left(\frac{\sqrt{s}}{\sigma}\right)\frac{1}{s} \\ &\times \left\{ -2\frac{\sqrt{s}}{\sigma}c + \frac{\xi^2\sqrt{s}}{\sigma^3}c^2 - \frac{\gamma\xi}{3\sigma}B\left(\frac{\sqrt{s}}{\sigma}\right)c \right\}. \end{aligned}$$

Since $(1/m) \sum_{n/m \geq a} h(n/m)$ is a Riemannian approximation to the integral $\int_0^\infty h(s)ds$, and converges to this limit as $m \rightarrow \infty$, we find in the end that $E\{Q_\varepsilon(c) - Q_\varepsilon(0)\}$ converges to

$$\begin{aligned} \lambda_a(c) &= \int_a^\infty \left[\frac{\xi^2}{\sigma^3} \frac{1}{\sqrt{s}} c^2 - \frac{2}{\sigma} \frac{1}{\sqrt{s}} c - \frac{\gamma}{3} \frac{\xi}{\sigma} \frac{1}{s} B\left(\frac{\sqrt{s}}{\sigma}\right) c \right] \phi\left(\frac{\sqrt{s}}{\sigma}\right) ds \\ &= 2 \int_{\sqrt{a}/\sigma}^\infty \left[\frac{\xi^2}{\sigma^2} c^2 - 2c - \frac{\gamma}{3} \frac{\xi}{\sigma} \frac{B(u)}{u} c \right] \phi(u) du. \end{aligned}$$

Some further analysis, using $B(u)/u = u^2 - 3$, finally proves (2.1). $\square$

In the applications presented below we will study this and similar limits as functions of c , to find the estimator sequence that can be expected to make fewest ε -errors. The optimal value of c as computed from (2.1) will depend upon the somewhat arbitrary value of a , however. Write in general $Q_{\varepsilon, m}(c)$ for the number of ε -errors committed by $(n/(n+c))\bar{X}_n$ among $n \geq m$ cases. The limit result (2.1) for $\lambda_a(c)$ relates to $Q_{\varepsilon, a/\varepsilon^2}(c)$, and it is natural to let a tend to zero. The limit function $\lambda_0(c)$ obtained by letting $a \rightarrow 0$ in (2.1) is not quite the limit of the expected difference between $Q_{\varepsilon, 1}(c)$ and $Q_{\varepsilon, 1}(0)$, since the remainder term $O(\sum_{n/\varepsilon \geq am} n^{-3/2})$ does not go to zero for $a = 1/m$. The following holds, however.

Proposition 1. Let $Q_\varepsilon(c)$ be the number of ε -misses for $(n/(n+c)) \bar{X}_n$, counted this time among $n \geq a(\varepsilon)/\varepsilon^2$, where $a(\varepsilon) \rightarrow 0$ while $a(\varepsilon)/\varepsilon^2 \rightarrow \infty$ (as with $a(\varepsilon) = \varepsilon$, for example). Then

$$\lambda_0(c) = \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon, a(\varepsilon)/\varepsilon^2(c) - Q_\varepsilon, a(\varepsilon)/\varepsilon^2(0)\} = \frac{\xi^2}{\sigma^2} c^2 - 2\left(1 - \frac{\gamma}{3} \frac{\xi}{\sigma}\right) c. \quad (2.7)$$

Sometimes it is reasonable to balance the objective $\bar{X}_n$ against some prior value. Bayesians often use estimators of the type $\hat{\xi}_n(c, d) = (n/(n+c)) \bar{X}_n + (c/(n+c)) \xi_0$; see Section 5. Consider therefore the slight extension

$$Q_\varepsilon(c, d) = \sum_{n \geq a/\varepsilon^2} I\left\{\left|\frac{n}{n+c} \bar{X}_n + \frac{c}{n+c} d - \xi\right| \geq \varepsilon\right\}.$$

The appropriate generalisations of (2.3) is found to be

$$l(c, d) = -\frac{1}{\sigma} \sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}} \frac{c(\xi - d)}{\sigma} - \frac{1}{\sqrt{nm}} \frac{c}{\sigma},$$

$$r(c, d) = \frac{1}{\sigma} \sqrt{\frac{n}{m}} + \frac{1}{\sqrt{n}} \frac{c(\xi - d)}{\sigma} + \frac{1}{\sqrt{nm}} \frac{c}{\sigma},$$

and by repeating previous arguments mutatis mutandis one proves the following generalisation of (2.7).

Proposition 2. Let as above ε and $a(\varepsilon) = \varepsilon$ tend to zero. Then

$$\lambda_0(c, d) = \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon(c, d) - Q_\varepsilon(0, 0)\} = \frac{(\xi - d)^2}{\sigma^2} c^2 - 2\left(1 - \frac{\gamma}{3} \frac{\xi - d}{\sigma}\right) c. \quad (2.8)$$

A further generalisation of (2.7) and (2.8) is given at the end of Section 4.3. We think of these as a.r.d. formulae; see (1.3) and the remarks made there. We also remark that there is an alternative way of proving (2.1), (2.7), (2.8), via distributional analysis following the Brownian motion approximation briefly discussed in Section 6. Details pertaining to this alternative proof are in the technical report of Hjort and Fenstad (1993) available from the authors.

Remark. One may also compute the Hodges and Lehmann asymptotic relative deficiency in this situation, defined as the limit of $n_0(c, d) - n_0$, where $n_0(c, d)$ is the sample size needed to achieve $E\{\hat{\xi}_n(c, d) - \xi\}^2 = E\{\hat{\xi}_{n_0}(0, 0) - \xi\}^2$, say; see Lehmann (1983, Section 5.2) for discussion. Then $n_0(c, d)/n_0(0, 0) \rightarrow 1$ for all (c, d) , but one can prove

$$\text{a.r.d.}_{\text{hl}}(c, d) = \lim_{n_0 \rightarrow \infty} \{n_0(c, d) - n_0(0, 0)\} = \frac{(\xi - d)^2}{\sigma^2} c^2 - 2c. \quad (2.9)$$

The a.r.d. formula (2.8) differs from this in that the skewness γ also enters, in a natural way.

3. Some applications

3.1. Normal mean

Let X_i be normal $(\theta, 1)$. Consider the estimator sequence $\hat{\theta}_n(c, d) = (n/(n+c))\bar{X}_n + (c/(n+c))d$, and let $Q_\varepsilon(c, d)$ be the number of ε -misses. This fits into the general scheme of Section 2 with $\xi = \theta$, $\sigma = 1$, and $\gamma = 0$. Hence

$$\lambda_0(c, d) = \lim_{\varepsilon \rightarrow 0} E_\theta \{Q_\varepsilon(c, d) - Q_\varepsilon(0, 0)\} = (\theta - d)^2 c^2 - 2c.$$

If this is averaged over θ w.r.t. some distributions with ‘prior mean’ $E\theta = \theta_0$ and ‘prior variance’ $\text{Var } \theta = \tau^2$, then one finds $\{\tau^2 + (\theta_0 - d)^2\}c^2 - 2c$, which is minimised when d is chosen as θ_0 and c is chosen as $1/\tau^2$. This provides fresh and independent motivation for using

$$\theta_n^* = \frac{n}{n + 1/\tau^2} \bar{X}_n + \frac{1/\tau^2}{n + 1/\tau^2} \theta_0 \quad (3.1)$$

in the presence of such prior knowledge, and agrees with both familiar Bayesian calculations in the normal model and the so-called credibility formula of actuarial statistics. Observe that θ_n^* can expect to make fewer ε -errors than $\bar{X}_n$ does if the true θ parameter is within $\sqrt{2}\tau$ of θ_0 . Note also that only θ_0 and τ^2 matter regarding the choice of prior weight function. Further discussion about risks and average risks is in Section 5.

3.2. Exponential mean

Let $X_i \sim \text{Exp}(1/\theta)$, and consider $Q_\varepsilon(c)$, the number of times $|\hat{\theta}_n(c)/\theta - 1| \geq \varepsilon$, where $\hat{\theta}_n(c) = (n/(n+c))\bar{X}_n$. This can be written $|(n/(n+c))\bar{Y}_n - 1| \geq \varepsilon$, where the Y_i ’s come from e^{-y} . This is as in Section 2 with $\xi = 1$, $\sigma = 1$, $\gamma = 2$. The limit of interest is $c^2 - \frac{2}{3}c$, with best value $c_0 = \frac{1}{3}$. The ML solution with $c = 0$ can be expected to make $\frac{1}{3}$ more ε -errors [sic] while the best estimator under squared error loss, which uses $c = 1$, can be expected to make $\frac{4}{3}$ more ε -errors.

3.3. Normal variance

Next consider $\hat{\sigma}_N^2(c) = \sum_{i=1}^N (Y_i - \bar{Y}_N)^2 / (N - 1 + c)$ for estimating the variance σ^2 based on data $Y_1, \dots, Y_N$ that are normal (μ, σ^2) . Study the number of times $|\hat{\sigma}_N^2(c)/\sigma^2 - 1| \geq \varepsilon$. This can be written $|(n/(n+c))(\chi_n^2/n) - 1| \geq \varepsilon$, writing n for $N - 1$, the degrees of freedom. This is once more as in Section 2, this time with χ_1^2 variables playing the role of X_i ’s. These have $\xi = 1$, $\sigma = \sqrt{2}$ and $\gamma = 2\sqrt{2}$. The limit to study is

$\frac{1}{2}c^2 - \frac{2}{3}c$, with minimum occurring for $c_0 = \frac{2}{3}$. Hence

$$\hat{\sigma}_N^2 = \frac{1}{N - (1/3)} \sum_{i=1}^N (Y_i - \bar{Y}_N)^2 \quad (3.2)$$

can expect to make the fewest ε -errors! See also Section 4.3.

3.4. Binomial probability

Let Y_n be binomial (n, p) , and let $Q_\varepsilon(c, d)$ count ε -misses for $(Y_n + cd)/(n + c)$. Note that Y_n is the sum of X_i 's with mean p , variance pq , and skewness $\gamma = (q - p)/(pq)^{1/2}$, where $q = 1 - p$. Hence the risk difference limit in (2.8) becomes

$$\text{risk}(c, d) = \frac{(p - d)^2}{pq} c^2 - 2c - \frac{2(p - q)(p - d)}{3pq} c.$$

After studying the case with 'prior guess' $d = \frac{1}{2}$ one can show that $p_n^* = (Y_n + 2/3)/(n + 4/3)$ is the second-order minimax sequence of estimators, in the sense of minimising the maximum possible risk. It can, seemingly, expect to make 2.667 fewer ε -errors than Y_n/n does, regardless of the underlying p !

There is an obstacle here, however, in that the distribution of X_i is lattice, the Cramér condition does not hold, and fine-tuned corrections are necessary for a formula like (2.6) to hold. Techniques from Kolassa and McCullagh (1990) are likely to provide in the end of a formula like (2.8) appropriate for the lattice case, but we have not pursued this. A possible trick to avoid these difficulties associated with exact a.r.d. calculations here is to spread the probability mass p for X_i uniformly on some tiny $[1 - \eta, 1 + \eta]$. Now results of Section 2 are in force, giving the best sequence of estimates for p . Letting $\eta \rightarrow 0$ in the end singles out the $(Y_n + 2/3)/(n + 4/3)$ sequence as the best. Note finally that the Beta $(2/3, 2/3)$ is a least favourable prior for p , in the sense of the calculations above.

4. Further examples

The aim of this section is to establish results in a couple of technically more demanding estimation problems. In each case somewhat strenuous modifications of the arguments used in Section 2 have to be devised in order to establish the limit of expected Q_ε -differences.

4.1. The squared mean in the normal model: known variance

Let X_i be normal (ξ, σ^2) , with σ known, and suppose $\theta = \xi^2$ is to be estimated. The ML solutions is $(\bar{X}_n)^2$, which could overestimate; the UMV solution is $(\bar{X}_n)^2 - \sigma^2/n$.

Study therefore $\hat{\theta}_n(d) = (\bar{X}_n) - d\sigma^2/n$ and its corresponding $Q_\varepsilon(d)$, the number of ε -misses among $n \geq a(\varepsilon)/\varepsilon^2$, with $a(\varepsilon)$ going to zero as in (2.7). We intend to show that

$$\lambda_0(d) = \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon(d) - Q_\varepsilon(0)\} = \left(\frac{1}{4}d^2 + \frac{1}{2}d\right)\frac{\sigma^2}{\xi^2}. \quad (4.1)$$

Some alterations to the programme of Section 2 are necessary. The indicator function summand in $Q_\varepsilon(d)$ is 1 if $|(\xi + T_n\sigma_n/\sqrt{n})^2 - d\sigma^2/n - \xi^2| \geq 1/\sqrt{m}$, in which $T_n = \sqrt{n}(\bar{X}_n - \xi)/\sigma$. This can be written

$$T_n^2 + 2(\xi/\sigma)\sqrt{n}T_n \geq d + \sqrt{n}\sqrt{s}/\sigma^2$$

or

$$T_n^2 + 2(\xi/\sigma)\sqrt{n}T_n \leq d - \sqrt{n}\sqrt{s}/\sigma^2,$$

where $s = n/m$ again. Let $b(d) < r(d)$ be the two roots of the first inequality and $c(d) < l(d)$ the two roots of the second inequality. Assume now that ξ is positive. Then $b(d) < c(d) < l(d) < 0 < r(d)$, and the indicator function term is

$$I\{T_n \leq b(d) \text{ or } T_n \geq r(d) \text{ or } c(d) \leq T_n \leq l(d)\}.$$

It turns out that both $b(d)$ and $c(d)$ are of size about $-2(\xi/\sigma)\sqrt{n}$, which is too far out on the left to be of significance, and we can concentrate on $I\{T_n \leq l(d) \text{ or } T_n \geq r(d)\}$. We need to compute the limit of

$$E\{Q_\varepsilon(d) - Q_\varepsilon(0)\} \doteq \sum_{n \geq am} [\{\Phi(l(d)) - \Phi(l(0))\} - \{\Phi(r(d)) - \Phi(r(0))\}], \quad (4.2)$$

cf. (2.4) and (2.5). A determined mind finds

$$\begin{aligned} l(d) &= -\sqrt{n}\frac{\xi}{\sigma} + \sqrt{n}\frac{\xi}{\sigma}\left(1 - \frac{1}{\sqrt{n}}\frac{\sqrt{s}}{\xi^2} + \frac{1}{n}\frac{d\sigma^2}{\xi^2}\right)^{1/2} \\ &\doteq -\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma} + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\frac{d\sigma}{\xi} - \frac{1}{8}\frac{s}{\xi^3\sigma}\right) + \frac{1}{n}\left(\frac{1}{4}d\sqrt{s}\frac{\sigma}{\xi^3} - \frac{1}{16}\frac{s^{3/2}}{\xi^5\sigma}\right) \\ &= -u + \delta_1(d)/\sqrt{n} + \delta_2(d)/n, \\ r(d) &= -\sqrt{n}\frac{\xi}{\sigma} + \sqrt{n}\frac{\xi}{\sigma}\left(1 + \frac{1}{\sqrt{n}}\frac{\sqrt{s}}{\xi^2} + \frac{1}{n}\frac{d\sigma^2}{\xi^2}\right)^{1/2} \\ &\doteq \frac{1}{2}\frac{\sqrt{s}}{\xi\sigma} + \frac{1}{\sqrt{n}}\left(\frac{1}{2}\frac{d\sigma}{\xi} - \frac{1}{8}\frac{s}{\xi^3\sigma}\right) + \frac{1}{n}\left(-\frac{1}{4}d\sqrt{s}\frac{\sigma}{\xi^3} + \frac{1}{16}\frac{s^{3/2}}{\xi^5\sigma}\right) \\ &= u + \delta_1(d)/\sqrt{n} - \delta_2(d)/n. \end{aligned}$$

One can now compute Taylor approximations to the terms of (4.2). The result, apart from $O(n^{-3/2})$ terms, is

$$\begin{aligned} & 2\phi(u)\{\delta_2(d) - \delta_2(0)\}/n + \{\tfrac{1}{2}\phi'(-u) - \tfrac{1}{2}\phi'(u)\}\{\delta_1(d)^2 - \delta_1(0)^2\}/n \\ &= \frac{1}{ms}\phi\left(\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\right)\left[\frac{1}{2}d\sqrt{s}\frac{\sigma}{\xi^3} + \left(\frac{1}{4}\frac{d^2\sigma^2}{\xi^2} - \frac{1}{8}\frac{ds}{\xi^4}\right)\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\right]. \end{aligned}$$

Further care and attention to details lead to (4.1).

When $\xi < 0$ then similar arguments give the same answer $(\frac{1}{4}d^2 + \frac{1}{2}d)\sigma^2/\xi^2$. We may conclude that the choice $d_0 = -1$ is best: Within the class of estimators under consideration, the estimator $(\bar{X}_n)^2 + \sigma^2/n$ has the second-order optimality property of making the fewest ε -errors, in expectation, in the limit as $\varepsilon \rightarrow 0$. The ML solution $(\bar{X}_n)^2$ can expect to make $\frac{1}{4}\sigma^2/\xi^2$ more ε -errors while the UMV $(\bar{X}_n)^2 - \sigma^2/n$ can expect to make σ^2/ξ^2 more ε -errors.

A natural extension is to include all estimators of the type $\hat{\theta}_n(c, d) = ((n/(n+c))\bar{X}_n)^2 - d\sigma^2/n$. Analysis is provided in Hjort and Fenstad (1993), giving the best possible values of c and d in terms of ‘prior averages’ of ξ^2/σ^2 and σ^2/ξ^2 ; see also Section 5. Included in Hjort and Fenstad (1993) is also analysis of the special case $\xi = 0$, where the asymptotics work differently.

4.2. The squared mean in the normal model: unknown variance

The extension to unknown σ is important because this parameter usually is unknown, of course, but also because the techniques given in a moment serve to illustrate how similar problems can be solved in other two- and multi-parameter situations.

The natural class of estimators to study is $\theta_n^*(d) = (\bar{X}_n)^2 - d\hat{\sigma}_n^2/n$, in which we take $\hat{\sigma}_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$ to be the unbiased version. Again $d = 0$ gives the ML solution while $d = 1$ gives the UMV estimator. Write $\hat{\sigma}_n^2 = \sigma^2 Z_n$, where Z_n is independent of $\bar{X}_n$ and distributed as $\chi_{n-1}^2/(n-1)$, so that in fact $\theta_n^*(d) = \hat{\theta}_n(dZ_n)$, in the notation of Section 4.1. The arguments there show that $Q_\varepsilon^*(d)$, the number of cases where $|\theta_n^*(d) - \xi^2| \geq \varepsilon$, essentially counts cases where $T_n \leq l(dZ_n)$ or $T_n \geq r(dZ_n)$. The analogue of (4.2) becomes

$$\begin{aligned} & \sum_{n \geq am} \{[\Pr\{T_n \leq l(dZ_n)\} - \Pr\{T_n \leq l(0)\}] \\ & - [\Pr\{T_n \leq r(dZ_n)\} - \Pr\{T_n \leq r(0)\}]\}. \end{aligned}$$

The simplest way to compute the limit of this sum, as $m = 1/\varepsilon^2$ grows, is to condition on the value of Z_n , use results of Section 4.1. and then integrate over the distribution $g_n(z_n)$ for Z_n . Carrying through this gives

$$\sum_{n \geq am} \int_0^\infty \frac{1}{ms}\phi\left(\frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\right)\left[\frac{1}{2}d\sqrt{s}z_n\frac{\sigma}{\xi^3} + \frac{1}{2}\frac{\sqrt{s}}{\xi\sigma}\left(\frac{1}{4}d^2z_n^2\frac{\sigma^2}{\xi^2} - \frac{1}{8}dz_n\frac{s}{\xi^4}\right)\right]g_n(z_n)dz_n.$$

But since $EZ_n = 1$ and $EZ_n^2 = 1 + 2/(n-1)$ the limit becomes in the end equal to the previous answer $(\frac{1}{4}d^2 + \frac{1}{2}d)\sigma^2/\xi^2$. We may conclude that $(\bar{X}_n)^2 + \hat{\sigma}_n^2/n$ is best.

It is interesting to compare this result to the corresponding one using the Hodges and Lehmann deficiency. Calculations with the mean squared error of $\theta_n^*(d)$ show that a.r.d.h., the limit of the sample size difference $n_0(d) - n_0(0)$, becomes $(\frac{1}{4}d^2 - \frac{1}{2}d)\sigma^2/\xi^2$, and the best estimator with this criterion is the UMV solution $(\bar{X}_n)^2 - \hat{\sigma}_n^2/n$.

4.3. Standard deviation in the normal model

Let us next present a variant of Section 3.3. Instead of counting instances of $|\hat{\sigma}_N^2(c)/\sigma^2 - 1| \geq \varepsilon$, penalise errors using the natural scale for σ and study

$$Q_\varepsilon(c) = \sum_{N \geq a/\varepsilon^2} I\{|\hat{\sigma}_N(c)/\sigma - 1| \geq \varepsilon\}. \quad (4.3)$$

The finer aspects of the second-order asymptotics machinery turn out to give a different answer than in Section 3.3 for the best value of c .

To find the best value this time we provide a modest but useful generalisation of (2.7) and (2.8). Suppose $h(\xi)$ is a smooth increasing transformation, and consider $Q_\varepsilon^*(c, d)$, the number of times $|h(\hat{\xi}_n(c, d)) - h(\xi)| \geq \varepsilon$, among $n \geq 1/\varepsilon$ (see (2.7)), with $\hat{\xi}_n(c, d) = (n\bar{X}_n + cd)/(n + c)$ as in Proposition 2 of Section 2. One can go through previous arguments, supplement them with further Taylor analysis, and prove

$$\begin{aligned} \lambda_0^*(c, d) &= \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon^*(c, d) - Q_\varepsilon^*(0, 0)\} \\ &= \frac{(\xi - d)^2}{\sigma^2} c^2 + \left\{ -2 + \frac{2\gamma}{3} \frac{\xi - d}{\sigma} + \frac{k_2}{k_1^2} (\xi - d) \right\} c, \end{aligned} \quad (4.4)$$

in which $k_1 = (h^{-1})'(h(\xi))$ and $k_2 = (h^{-1})''(h(\xi))$. The necessary details are furnished in Hjort and Fenstad (1993). Elementary rules show further that k_2/k_1^2 also can be written as $-h''(\xi)/h'(\xi)$. The special case $h(\xi) = \xi$ has $k_2 = 0$ and gives back (2.8).

To analyse the problem associated with (4.3) we can put $h(\xi) = \xi^{1/2}$, $d = 0$, and use $\xi = 1$ in the end. This gives

$$\lambda_0(c) = \lim_{\varepsilon \rightarrow 0} E\{Q_\varepsilon(c) - Q_\varepsilon(0)\} = c^2/\tau^2 + ((2/3)\gamma/\tau - 3/2)c,$$

in terms of standard deviation τ and skewness γ . In the particular χ_1^2 case we find $\lambda_0(c) = \frac{1}{2}c^2 - \frac{1}{6}c$, with minimum occurring for $c_0 = \frac{1}{6}$. We may conclude that the estimator sequence $\hat{\sigma}_N = \{(N - (5/6))^{-1} \sum_{i=1}^N (Y_i - \bar{Y}_N)^2\}^{1/2}$ can be expected to make fewest errors of the type $|\hat{\sigma}_N/\sigma - 1| \geq \varepsilon$. Let us also use (4.4) to exhibit the estimator sequence that can be expected to produce the fewest ε -errors on log scale. So let $Q_\varepsilon^*(c)$ count instances of $|\log \hat{\sigma}_N^2(c) - \log \sigma^2| \geq \varepsilon$. Then the limit can be shown to be $\lambda_0^*(c) = \frac{1}{2}c^2 - (\frac{2}{3} - e^{-1})c$, with minimum for $c_0 = \frac{2}{3} - e^{-1}$. So using denominator

$N - 1 + c_0$, i.e. $\sigma_N^* = \{(N - 0.695)^{-1} \sum_{i=1}^N (Y_i - \bar{Y}_N)^2\}^{1/2}$, can be expected to make fewest errors for σ^p on the log scale, for every value of p .

Remark. Formula (3.2) with $N - 1/3$ and the two recently developed $N - 5/6$ and $N - 0.695$ add three new items to the distinguished list of denominators in the variance estimation formula $\sum_{i=1}^N (Y_i - \bar{Y}_n)^2 / (N - 1 + c)$ reached by various statistical principles. This list includes: (i) The maximum likelihood estimator, normal-based or nonparametric, has $1/N$. (ii) The unbiased estimator for σ^2 with smallest variance, under normality or under nonparametric circumstances, has $1/(N - 1)$. (iii) If one wants $E\hat{\sigma}_N(c) = \sigma$, unbiasedness on natural scale, then one needs $E(\chi_n^2)^{1/2}/(n + c)^{1/2} = 1$, using $n = N - 1$ again, whose approximate solution is $c = -\frac{1}{2}$, leading to $1/(N - \frac{3}{2})$. (iv) Another possibility is to minimise mean squared error $E_\sigma(\hat{\sigma}_N^2 - \sigma^2)^2$. The best constant is $1/(N + 1)$. This also gives the best invariant estimator under squared error loss for σ^2 . (v) It appears natural to solve this problem using natural scale, i.e. to minimise $E_\sigma(\hat{\sigma}_N - \sigma)^2$. This can again be done uniformly in σ , and the solution is $c = n^2/\{E(\chi_n^2)^{1/2}\}^2 - n$, for which $\frac{1}{2}$ is an approximation. The best denominator is $N - \frac{1}{2}$, and gives at the same time the best invariant estimator under squared error loss for σ . (vi) The median of χ_n^2 is $n - 2/3 + (4/27)/n$ to a good approximation, by the Wilson–Hilferty formula. Hence denominator $N - 5/3 + 4/27N \doteq N - 5/3$ gives an approximate median-unbiased estimator; it overshoots as often as it undershoots. (vii) Since $E \log(\chi_n^2) = \log 2 + \psi(n/2) \doteq \log n - (1/n + 1/3n^2)$ one finds that $\tilde{\sigma}$ with denominator $N - 2 + 1/6N \doteq N - 2$ gives unbiasedness on log scale. This is a good property both because the log of the scale is a natural quantity and because $\tilde{\sigma}^p$ becomes log-unbiased for σ^p for every value of p . (viii) The Bayes solution under a vague prior (log σ uniform on the line) uses $N - 3$ if loss is squared error on σ^2 scale and $N - 1$ if loss is squared error on $1/\sigma^2$ scale.

5. Loss functions and Bayes solutions

5.1. Decision theoretic framework

Who can understand his errors (Psalm 19: 12)? And who can count them? We are essentially working with the somewhat nonstandard loss function L_ε that for a given sequence of estimates $\{\hat{\theta}_n\}$ counts

$$L_\varepsilon[\theta, \{\hat{\theta}_n: n \geq 1\}] = \sum_{n \geq a(\varepsilon)/\varepsilon^2} I\{|\hat{\theta}_n - \theta| \geq \varepsilon\}, \quad (5.1)$$

the number of ε -errors among all $n \geq a(\varepsilon)/\varepsilon^2$ cases, for a very small ε . The associated risk function is

$$R_\varepsilon(\theta) = E_\theta L_\varepsilon[\theta, \{\hat{\theta}_n: n \geq 1\}] = \sum_{n \geq a(\varepsilon)/\varepsilon^2} \Pr_\theta\{|\hat{\theta}_n - \theta| \geq \varepsilon\}. \quad (5.2)$$

In some cases this series diverges while the difference risk $R_{1,\varepsilon}(\theta) - R_{2,\varepsilon}(\theta)$ (say) might constitute a convergent series. The $\lambda_a(c)$ and $\lambda_0(c)$ functions found in Section 2 are indeed limits of $R_{c,\varepsilon}(\theta) - R_{0,\varepsilon}(\theta)$ as $\varepsilon \rightarrow 0$, where $R_{c,\varepsilon}$ refers to the risk function for the $\hat{\theta}_n(c)$ sequence. One might next obtain average risk with respect to a weight function over the parameter space, which one may choose or not choose to interpret as a prior distribution in the Bayesian fashion, and finally minimise this expression w.r.t. the class of estimators under consideration.

This was the programme partially carried out in Sections 3.1 and 4.1. A more direct and general approach which avoids restriction to a given class of estimators is possible, as follows, provided the prior is *fully* specified. Let $\pi(\cdot)$ be the prior distribution and $\pi_n(\cdot) = \pi_n(\cdot|x_1, \dots, x_n)$ the posterior distribution at step n . Then

$$\int R_\varepsilon(\theta)\pi(\theta) d\theta = \sum_{n \geq a/\varepsilon^2} \int \cdots \int \pi_n\{|\theta - \hat{\theta}_n| \geq \varepsilon\} f_n(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

where f_n is the marginal density resulting from having $\pi(\cdot)$ as prior. For given ε the best solution is to minimise each term, which means choosing $\hat{\theta}_n$ to maximise $\pi_n\{|\theta - \hat{\theta}_n| \leq \varepsilon\}$, for each given $x_1, \dots, x_n$. When ε goes to zero for fixed n this would mean using the posterior mode. The balance is more delicate here, however, where summing over n comes before letting ε tend to zero. Think of $\hat{\theta}_n$ as $\theta_n^* + u/\sqrt{n}$, where θ_n^* is the posterior mean for θ . Thus one needs to minimise $\pi_n\{|\sqrt{n}(\theta - \theta_n^*) - u| \geq \sqrt{s}\}$ w.r.t. u , where $m = 1/\varepsilon^2$ and $s = n/m$ as in earlier sections. A full analysis would call for Edgeworth expansion again. Observe in particular that the skewness of the posterior distribution plays a role here.

5.2. A normal mean

Consider once more the situation of Section 3.1, where it was noted that θ_n^* of (3.1) was best among all linear functions of the sample mean, for any prior weight function with first moment θ_0 and variance τ^2 . Suppose now that θ is given the normal prior with these parameters. Then $\pi_n(\cdot)$ is also normal, with mean equal to θ_n^* of (3.1) and variance $\tau_n^2 = \tau^2/(n\tau^2 + 1)$. The symmetry of the normal density shows that $\pi_n\{|\theta - \hat{\theta}_n| \leq \varepsilon\}$ is maximised for $\hat{\theta}_n = \theta_n^*$. So (3.1) is not only the best solution within the linear class $\hat{\theta}_n(c, d)$ but the very best of *all* estimators, under the average EQ_ε criterion, provided the prior used is indeed $N\{\theta_0, \tau^2\}$. This even holds for each positive ε as well as in the limit.

6. Second-order limits in distribution

The starting point for the present work was the first-order result (1.1) of Hjort and Fenstad (1992), relating the distribution of $\varepsilon^2 Q_\varepsilon$ to the amount of time Brownian

motion $W(s)$ spends outside the $|W(s)| \leq s/\sigma$ area. For two estimator sequences with the same limiting distribution the typical situation is that $\varepsilon^2(Q_{1,\varepsilon} - Q_{2,\varepsilon}) \rightarrow_p 0$ and $Q_{1,\varepsilon}/Q_{2,\varepsilon} \rightarrow_p 1$, and in the course of the paper we have provided limit formulae for the mean value of $Q_{1,\varepsilon} - Q_{2,\varepsilon}$ in different situations. This final section briefly presents some second-order *distributional* aspects of this difference, thus providing additional information about what happens when $\varepsilon \rightarrow 0$. In several situations this distribution is in the limit related to the amount of time Brownian motion spends along the two border lines $\pm s/\sigma$. Such distributions, for Brownian motion and for general diffusion processes, are studied in Hjort and Khasminskii (1993).

To illustrate the typical behaviour, let us go back to the situation of Section 2. Take c and ξ to be positive so that $l(c) > l(0)$ and $r(c) > r(0)$ in (2.3). It follows from (2.4) that

$$Q_\varepsilon(c) - Q_\varepsilon(0) = \sum_{n \geq a/\varepsilon^2} [I\{l(0) \leq T_n \leq l(c)\} - I\{r(0) \leq T_n \leq r(c)\}]. \quad (6.1)$$

This is an infinite sum of quite rare 1's and -1 's. Our previous labours were confined to the expected value of this difference and therefore stayed within the realm of probabilities and Edgeworth expansions. Careful analysis of a different type is called for to find the limit distribution of the (6.1) difference. Details are provided partly in Hjort and Khasminskii (1993, Section 6.3) and more fully in the technical report of Hjort and Fenstad (1993). The bottom line is that

$$\varepsilon\{Q_\varepsilon(c) - Q_\varepsilon(0)\} \rightarrow_d A - B, \quad (6.2)$$

say, where A and B are two variables concerned respectively with how much time Brownian motion spends along s/σ and $-s/\sigma$. In the case of a fixed positive a in (6.1) each of A and B is a mixture of an exponential and a unit point mass at zero, and the two are dependent. In particular $A - B$ can then be zero with positive probability. In the case of a shrinking $a = a(\varepsilon)$, say $a = \varepsilon$ in (6.1), both A and B are exponentially distributed with mean $c\xi$ and with intercorrelation $-\frac{1}{3}$.

References

- Barndorff-Nielsen, O.E. and D.R. Cox (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.
- Berkson, J. (1980). Minimum chi-square, not maximum likelihood! (with discussion). *Ann. Statist.* **8**, 457–487.
- Ghosh, J.K. and K. Subramanyam (1974). Second order efficiency of maximum likelihood estimators. *Sankhyā Ser. A* **36**, 325–258.
- Hjort, N.L. and G. Fenstad (1992). On the last time and the number of times an estimator is more than ε from its target value. *Ann. Statist.* **20**, 469–489.

- Hjort, N.L. and G. Fenstad (1993). Second order asymptotics for the number of times an estimator is more than ε from its target value. Statistical Research Report, Department of Mathematics, University of Oslo.
- Hjort, N.L. and R.Z. Khasminskii (1993). On the time a diffusion process spends along a line. *Stochastic Process. Appl.* **47**, 229–247.
- Hodges, J. and E.L. Lehmann (1970). Deficiency. *Ann. Math. Statist.* **41**, 783–801.
- Kolassa, J.E. and P. McCullagh (1990). Edgeworth series for lattice distributions. *Ann. Statist.* **18**, 981–985.
- Lehmann, E.L. (1983). *Theory of Point Estimation*. Wiley, New York.
- Pfanzagl, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1**, 993–1026.
- Rao, C.R. (1962). Efficient estimates and optimum inference procedures in large samples (with discussion). *J. Roy. Statist. Soc. Ser. B* **24**, 46–72.