CWI Quarterly Volume 9 (1 & 2) 1996, pp. 69 – 76

Estimation in Restricted Parameter Spaces – Some History and Some Recent Developments

C. van Eeden[1]

Department of Statistics
The University of British Columbia
Vancouver V6T 1Z2 Canada

Département de mathématiques
Université du Québec à Montréal
Montréal H3C 3P8 Canada

This paper traces the development of restricted-parameter-space estimation problems from the early 50's, when the emphasis was on maximum likelihood, to the present when questions of admissibility and minimaxity of these and related estimators are being addressed. Open problems are pointed out.

1. Introduction

This paper is based on an invited talk presented at the SMC50 Mathematics Congress which was held on the occasion of CWI's 50th anniversary in February 1996. It is concerned with problems of estimation in restricted parameter spaces, in particular with some of this problem's history and development.

The problem arose, in the early 1950's, out of a practical problem in which two probabilities θ_1 and θ_2, known to satisfy $\theta_1 \leq \theta_2$, needed to be estimated. Maximum likelihood estimation was used for this purpose. Later, maximum likelihood estimators (MLEs) were shown to be inadmissible for squared error loss. That is, it was shown that there exist estimators which are better than the MLE in the sense that their expected loss, as a function of the parameter to be estimated, is nowhere larger and somewhere smaller than that of the MLE. This then led to the search for dominators for these inadmissible estimators, as well as for admissible estimators with "good" properties. One such property is

[1] Address for correspondence : Moerland 19, 1151 BH Broek in Waterland, The Netherlands; e-mail: cve@xs4all.nl

that of minimaxity, where an estimator is minimax when there does not exist an estimator with a smaller maximum expected loss.

In Section 2, the early results on MLE are sketched. Section 3 contains results on admissibility, for squared error loss, of the MLE and related classes of estimators. Results on minimaxity, for (scale-invariant) squared error loss, are given in Section 4. Other questions concerning estimation problems in restricted parameter spaces, as well as some remarks on methods of proof, can be found in Section 5. Open problems are mentioned throughout the paper.

Given the restrictions on time (a half-hour talk) and space in this journal, it is not possible to give a complete account of the literature on the subject, nor of the various open problems. But the quoted references together with their references should provide a good coverage of the subject.

2. Maximum likelihood estimators

Sometime in the early 1950's, somebody came for advice to the Statistical Consultation Service of the Mathematical Center in Amsterdam with a practical problem which led to the following question. Suppose we have two independent random variables X_1 and X_2, where, for $i = 1, 2$, X_i is $\text{Bin}(n_i, \theta_i)$ and suppose we know that $\theta_1 \leq \theta_2$. How does one estimate $\theta = (\theta_1, \theta_2)$? The client was given the maximum likelihood estimator $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$, where $\hat{\theta}_i = X_i/n_i, i = 1, 2$ if $X_1/n_1 \leq X_2/n_2$ and $\hat{\theta}_1 = \hat{\theta}_2 = (X_1 + X_2)/(n_1 + n_2)$ if not.

Professor Hemelrijk, at the time Head of the Statistical Consultation Service at the Mathematical Center, then asked me to look at possible generalizations of this problem. This led to the study of the k-sample problem, where $X_{i,j}, j = 1, \ldots, n_i, i = 1, \ldots, k, k \geq 2$ are independent random variables and the $X_{i,j}, j = 1, \ldots, n_i$ have distribution function $F(x; \theta_i)$, $\theta_i \in R, i = 1, \ldots, k$. The parameter space Θ for $\theta = (\theta_1, \ldots, \theta_k)$ was determined by inequalities among the θ_i, either by a complete ordering, i.e. $\Theta = \{\theta | \theta_1 \leq \ldots \leq \theta_k\}$, or an incomplete one such as, e.g., $k = 3$ and $\Theta = \{\theta | \theta_1 \leq \theta_3, \theta_2 \leq \theta_3\}$. The question posed was to find the MLE, $\hat{\theta}$, of θ. For definiteness, note that, here and in the rest of this paper, estimators of a vector θ which is restricted to Θ are functions from the sample space into the parameter space Θ.

Conditions for the existence of the MLE, as well as algorithms for finding it can be found in van Eeden [34, 35, 36, 37, 38]. Examples of distributions for the $X_{i,j}$ where these conditions are satisfied are the one-parameter exponential family and the uniform distribution on the interval $(0, \theta_i)$. The algorithm for the completely ordered case is what later became to be known as the pool-adjacent-violators algorithm (PAVA). It says that $\hat{\theta}_i = \hat{\theta}_{i+1}$ when $t_i > t_{i+1}$, where $t = (t_1, \ldots, t_k)$ is the unrestricted MLE of θ. For the incompletely ordered case the algorithms are generalizations of PAVA or are PAVA-like algorithms.

The completely ordered binomial case was also solved by Ayer, Brunk, Ewing, Reid and Silverman [1]. Further, Brunk [9, 10] considered the k-sample problem where the $X_{i,j}$ have a one-parameter exponential family distribution. These authors give the PAVA for the completely ordered case and an explicit formula for the general case.

Many of these as well as other results on MLEs for ordered parameters can be found in BARLOW, BARTHOLOMEW, BREMNER and BRUNK [2] and in ROBERTSON, WRIGHT and DYKSTRA [26]. Each of these books also contains many results on problems of tests of hypotheses concerning ordered parameters.

3. QUESTIONS OF ADMISSIBILITY

In the early days of the development of the subject of restricted-parameter-space estimation, there does not seem to have been much interest, if any, in the properties of the MLE. It seems that it was not known at that point in time that estimators which have "good" properties in unrestricted parameter spaces lose many of these properties when the parameter space is restricted. As an example, for X a $\mathcal{N}(\theta, 1)$ random variable and squared error loss, the MLE of θ for the parameter space $\{\theta | -\infty < \theta < \infty\}$ is unbiased, admissible, minimax and it has a normal distribution. For the parameter space $\{\theta | \theta \geq 0\}$ the MLE is biased and inadmissible, but still minimax. It does not have a normal distribution and for $\theta = 0$ it is, for a sample $X_1, \ldots, X_n$, not even asymptotically normal. On the other hand, there are examples where the MLE does not lose its admissibility property when the parameter space is restricted. E.g., when X is $\mathrm{Bin}(n, \theta)$ with $\theta \in [0, a]$ for some known $a \in (0, 1)$ and $n \geq 2, 1 \leq an \leq 2$, the MLE of θ is admissible for squared error loss (see CHARRAS [12] and CHARRAS and VAN EEDEN [13]).

The above quoted inadmissibility result for the lower bounded normal mean case was (as far as I know) first proved by SACKS [29]. He obtained this result from his necessary conditions for admissibility of estimators of the canonical parameter in a one-parameter exponential family distribution. However, it was not until the middle 1970's that admissibility started to be studied in a more systematic way. For instance, SACKROWITZ and STRAWDERMAN [28] show that, for unweighted squared error loss, the MLE of completely ordered θ_i, when X_i is $\mathrm{Bin}(n_i, \theta_i), i = 1, \ldots, k$, is admissible if and only if either $\sum_{i=1}^{k} n_i < 7$; or $k = 2$ and $n_1 = 1$ or $n_2 = 1$; or $k = 3$ and $n_1 = n_3 = 1$.

This result got me interested in the problem again which led to the PhD thesis in 1979 of CHARRAS [12] and several papers (CHARRAS and VAN EEDEN [13, 14, 15, 16]) based on it. The results in these publications deal with the general problem of a probabilty space $(\mathcal{X}, \mathcal{A}, P_\theta)$, where $\theta \in \Theta$ and Θ is a closed, convex subset of R^k. Further, the family $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ is dominated by a σ-finite measure. Conditions are given under which "boundary estimators" are inadmissible for squared error loss, where a "boundary estimator" is an estimator which takes, with positive probability for some $\theta \in \Theta$, values on or "near" the boundary of Θ. In order to obtain their results, these authors start with sufficient conditions for boundary estimators not to be (limits of) Bayes estimators. They then add sufficient conditions for estimators which are not (limits of) Bayes to be inadmissible. For some cases these authors also give dominators for inadmissible MLEs.

MOORS [24, 25] considers the same set-up. He assumes the problem to be invariant with respect to a certain group of transformations and obtains inadmissibility results by constructing dominators for boundary estimators (terminology introduced by Moors).

For the one-parameter exponential family, BROWN [8] gives necessary conditions for estimators of the expectation parameter to be admissible for squared error loss.

These three sets of results jointly cover many cases of inadmissible MLEs and other boundary estimators for restricted parameter spaces.

As already mentioned above, dominators for inadmissible boundary estimators were obtained by Charras and van Eeden, as well as by Moors. Other cases where dominators are known are the case of completely ordered binomial probabilities, where SACKROWITZ [27] gives dominators for unweighted squared error loss for the (inadmissible) MLE. SHAO and STRAWDERMAN [32, 33] give dominators for (inadmissible) truncated linear estimators for the lower bounded normal mean and for the lower bounded scale parameter of a gamma distribution. They use squared error loss. However, for many cases dominators for inadmissible boundary estimators are not known. As an example, I do not know any dominator, for squared error loss, for the MLE of a lower bounded mean of a Poisson distribution. Further, the known dominators all seem to be inadmissible, leaving open the problem of finding admissible dominators.

4. ADMISSIBLE MINIMAX ESTIMATORS

The first example I know of of an admissible minimax estimator for a restricted parameter space is KATZ's [21] estimator for the lower bounded mean of a normal distribution for squared error loss. His paper addresses the more general problem of estimation of the lower bounded expectation parameter of an exponential family, but his admissibility proof for the general case is incorrect (see VAN EEDEN [39]).

Another case where an admissible minimax estimator for KATZ's problem is known is the case of a lower bounded scale parameter θ of a gamma distribution for scale-invariant squared error loss. For $\theta \geq a > 0$, VAN EEDEN [39] obtains this estimator as the pointwise limit of a sequence of Bayes estimators. Further, an admissible minimax estimator for θ where X has density $\exp(-(x-\theta)), x \geq \theta, \theta \geq 0$ was obtained by BERRY [3] for squared error loss.

These three cases are the only ones I know of where an admissible minimax estimator is known for the restricted parameter space $[a, \infty)$. I have been working on the case of the mean θ of a Poisson distribution with $\theta \geq a > 0$ and loss function $L(d, \theta) = (\theta - d)^2/\theta$, but have not succeeded in finding an admissible minimax estimator. However, the following reasoning gives the minimax risk for this case. Let $\delta(x) = \max(a, x)$. Then it is easily seen that $\sup_{\theta \geq a} \mathcal{E}_\theta L(\delta(X), \theta) = 1$, which gives an upper bound of 1 for the minimax risk. Using the information inequality one can prove (as LEHMANN [22, pp.

267–268], does for the case of the normal mean) that the minimax risk for this problem is ≥ 1, showing at the same time that the estimator δ is minimax. This technique of obtaining the minimax risk for the parameter space $[a, \infty)$ might well be applicable in other cases. Lower bounds such as, e.g., those obtained by GAJEK and KAŁUSZKA [19], GILL and LEVIT [20], and SATO and AKAHIRA [30, 31], could be helpful here.

For the case of the parameter space $[a, b], -\infty < a < b < \infty$, CASELLA and STRAWDERMAN [11], as well as ZINZIUS [41], show that, for $X\ \mathcal{N}(\theta, 1)$ and squared error loss, there exists a unique admissible minimax estimator of θ. This estimator is Bayes with respect to a prior with support $\{a, b\}$.

These results led to a sequence of papers on the problem where $X_1, \ldots, X_n$ are independent, identically distributed with density $f(x-\theta)$ or $f(x/\theta)/\theta$ where $\theta \in [a, b]$ and the loss function $|d - \theta|^p$ with $p > 1$. These cases were considered by Bishoff, Chen, Eichenauer-Herrmann, Fieger, Ickstadt, Ochtrop and Wulfert. References to their papers can be found, e.g., in BISCHOFF and FIEGER [5], BISCHOFF, FIEGER and OCHTROP [6] and BISCHOFF, FIEGER and WULFERT [7]. The results of these authors are similar to those of Casella, Strawderman and Zinzius : under certain regularity conditions on the density of the X_i and for small enough $b - a$, there exists a unique admissible minimax estimator and this estimator is Bayes with respect to a prior with support $\{a, b\}$. It should be noted here that, for the case where $p = 1$, there does not necessarily exist such a prior (see EICHENAUER-HERRMANN and ICKSTADT [17]).

The case where the X_i have density $f(x/\theta)/\theta$ with $\theta \in [a, b], a > 0$, and scale-invariant squared error loss, is considered by VAN EEDEN and ZIDEK [40]. They obtain the "same" result again, except that now $(b/a) - 1$ needs to be small.

Not much is known for larger values of $b - a$, except that the number of points in the support of the prior increases as $b - a$ increases. Approximations to the minimax risk, as well as estimators which are approximately minimax for $b - a$ large were obtained by BICKEL [4] and by LEVIT [23] for the normal mean case with squared error loss. Similar results for the scale problem with scale-invariant squared error loss and $\log \theta \in [a, b]$ can be found in GAJEK and KAŁUSZKA [18].

5. SOME REMARKS AND SOME MORE RESULTS

Nothing was said above about why boundary estimators are, in general, not admissible. As an example of a proof, look at the case where X is $\mathcal{N}(\theta, 1)$ with $\theta \in [a, b]$. The MLE, $\hat{\theta}$, is given by $\hat{\theta}(x) = x$ when $x \in [a, b]$, $= a$ when $x < a$ and $= b$ when $x > b$. To prove inadmissibilty, for squared error loss, it is sufficient to prove that $\hat{\theta}$ is not Bayes. In order to prove this suppose that $\hat{\theta}$ is Bayes with respect to some prior on $[a, b]$. Then, because $\hat{\theta}$ is the mean of the posterior distribution, the support of the posterior is $\{a\}$ when $x < a$ and $\{b\}$ when $x > b$. The contradiction then comes from the fact that, for this case, the prior and the posterior have, for all x and all priors, the same support. This reasoning, which is the one used by CHARRAS [12] and CHARRAS and VAN EEDEN [13], does not work for the example given above where X is $\text{Bin}(n, \theta)$ with $\theta \in [0, a]$ for some $a \in (0, 1)$, because for this case the prior and the posterior do not necessarily have the same support.

About how to find (admissible) minimax estimators. The problem is more complicated than for unrestricted parameter spaces where, in general, finding a Bayes estimator with a constant risk function solves the problem. For restricted parameter spaces the risk function of a minimax estimator is, in general, not constant. What one can try is use the result (see, e.g., LEHMANN [22], p. 256) that, when the sequence of Bayes risks of a sequence of Bayes estimators has a limit, $r < \infty$ say, and δ is an estimator whose risk function is upper bounded by r, then δ is minimax. This result was used by KATZ [21] for his normal mean result and by VAN EEDEN [39] for her gamma scale result. In each of these two cases the minimax estimator is the pointwise limit of the sequence of Bayes estimators.

As a final remark, there are several problems concerning estimation in restricted parameter spaces which I have not, or barely, touched upon in the above but for which results have been obtained. To name a few of them : i) minimax estimation of a vector $\theta = (\theta_1, \ldots, \theta_k)$ with $k > 1$; ii) loss functions other than (scale-invariant) squared error; iii) the search for minimax estimators within restricted classes of estimators, such as, e.g., truncated linear ones - this might simplify the search without increasing the minimax risk too much; iv) problems with nuissance parameters, such as, e.g., the problem of estimating θ_1 when, for $i = 1, 2$, X_i is Poisson with mean θ_i, $\theta_1 \leq \theta_2$ and the X_i are independent; v) approximations to minimax estimators, either for restricted parameter spaces which are "almost unrestricted", or for large sample sizes.

REFERENCES

1. M. AYER, H.D. BRUNK, G.M. EWING, W.T. REID and E. SILVERMAN (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**, pp. 641–647.
2. R.E. BARLOW, D.J. BARTHOLOMEW, J.M. BREMNER and H.D. BRUNK (1972). *Statistical Inference under Order Restrictions. The Theory and Application of Isotonic Regression.*Wiley, New York.
3. J.C. BERRY (1993). Minimax estimation of a restricted exponential location parameter. *Statist. Decisions* **11**, pp. 307–316.
4. P.J. BICKEL (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.* **9**, pp. 1301–1309.
5. W. BISCHOFF and W. FIEGER (1993). On least favourable two point priors and minimax estimators under absolute error loss. *Metrika* **40**, pp. 283–298.

6. W. BISCHOFF, W. FIEGER and S. OCHTROP (1995). Minimax estimation for the bounded mean of a bivariate normal distribution. *Metrika* **42**, pp. 379–394.
7. W. BISCHOFF, W. FIEGER and S. WULFERT (1995). Minimax- and Γ-minimax estimation of a bounded normal mean under linex-loss. *Statist. Decisions* **13**, pp. 287–298.
8. L.D. BROWN (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* Lecture Notes-Monograph Series **9**, Institute of Mathematical Statistics, Hayward, California.
9. H.D. BRUNK (1955). Maximum likelihood estimates of monotone parameters. *Ann. Math. Statist.* **26**, pp. 607–616.
10. H.D. BRUNK (1958). On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29**, pp. 437–454.
11. G. CASELLA and W.E. STRAWDERMAN (1981). Estimating a bounded normal mean. *Ann. Statist.* **9**, pp. 870–878.
12. A. CHARRAS (1979). *Propriété Bayesienne et Admissibilité d'Estimateurs dans un Sous-ensemble Convex de R^p.* Ph.D. thesis, Université de Montréal.
13. A. CHARRAS and C. VAN EEDEN (1991a). Bayes and admissibility properties of estimators in truncated parameter spaces. *Canad. J. Statist.* **19**, pp. 121–134.
14. A. CHARRAS and C. VAN EEDEN (1991b). Limits of Bayes estimators in convex truncated parameter spaces. *Statist. Probab. Lett.* **11**, pp. 479–483.
15. A. CHARRAS and C. VAN EEDEN (1992). Bayes properties of estimators of location parameters in truncated parameter spaces. *Statist. Decisions* **10**, pp. 81–86.
16. A. CHARRAS and C. VAN EEDEN (1994). Inadmissibility for squared loss when the parameter to be estimated is restricted to the interval $[a, \infty)$. *Statist. Decisions.* **12**, pp. 257–266.
17. J. EICHENAUER-HERRMANN and K. ICKSTADT (1992). Minimax estimation for a bounded location parameter. *Metrika* **39**, pp. 227–237.
18. L. GAJEK and M. KAŁUSZKA (1994). Lower bounds for the asymptotic Bayes risk in the scale model (with applications to the second order minimax estimation). *Ann. Statist.* **22**, pp. 1831–1839.
19. L. GAJEK and M. KAŁUSZKA (1995). Nonexponential applications of a global Cramér-Rao inequality. *Statistics* **26**, pp. 111–122.
20. R.D. GILL and B. Y. LEVIT (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli* **1**, pp. 59–79.
21. M.W. KATZ (1961). Admissible and minimax estimators of parameters in truncated parameter spaces. *Ann. Math. Statist.* **32**, pp. 136–142.
22. E.L. LEHMANN (1983). *Theory of Point Estimation.* Wiley, New York.
23. B.Ya. LEVIT (1980). On asymptotic minimax estimates of second order. *Theory Probab. Appl.* **25**, pp. 552–568.
24. J.J.A. MOORS (1981). Inadmissibility of linearly invariant estimators in truncated parameter spaces. *J. Amer. Statist. Assoc.* **76**, pp. 910–915.

25. J.J.A. MOORS (1985). *Estimation in Truncated Parameter Spaces.* Ph.D. thesis, Tilburg University, Tilburg, The Netherlands.
26. T. ROBERTSON, F.T. WRIGHT and R.L. DYKSTRA (1988). *Order Restricted Statistical Inference.*Wiley, New York.
27. H. SACKROWITZ (1982). Procedures for improving the MLE for ordered binomial parameters. *J. Statist. Plann. Inference* **6**, pp. 287–296.
28. H. SACKROWITZ and W. STRAWDERMAN (1974). On the admissibility of the MLE for ordered binomial parameters. *Ann. Statist.* **2**, pp. 822–828.
29. J. SACKS (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist.* **34**, pp. 751–768.
30. M. SATO and M. AKAHIRA (1995). Information inequalities for the minimax risk. *J. Japan Statist. Soc.* **25**, pp. 151–158.
31. M. SATO and M. AKAHIRA (to appear). An information inequality for the Bayes risk. *Ann. Statist.*
32. P.Y-S. SHAO and W.E. STRAWDERMAN (to appear). Improving on truncated linear estimates of a gamma scale parameter. *Canad. J. Statist.*
33. P.Y-S. SHAO and W.E. STRAWDERMAN (1996). Improving on the MLE of a positive normal mean. *Statist. Sinica* **6** , pp. 259–274.
34. C. VAN EEDEN (1956). Maximum likelihood estimation of ordered probabilities. *Proc. Kon. Nederl. Akad. Wetensch.* A **59**, pp. 444–455.
35. C. VAN EEDEN (1957a). Maximum likelihood estimation of partially or completely ordered parameters. *Proc. Kon. Nederl. Akad. Wetensch.* A **60**, pp. 128–136 and 201–211.
36. C. VAN EEDEN (1957b). Note on two methods for estimating ordered parameters of probability distributions. *Proc. Kon. Nederl. Akad. Wetensch.* A **60**, pp. 506–512.
37. C. VAN EEDEN (1957c). A least squares inequality for maximum likelihood estimates of ordered parameters. *Proc. Kon. Nederl. Akad. Wetensch.* A **60**, pp. 513–521.
38. C. VAN EEDEN (1958). *Testing and Estimating Ordered Parameters of Probability Distributions.* Ph.D. thesis, University of Amsterdam.
39. C. VAN EEDEN (1995). Minimax estimation of a lowerbounded scale parameter of a gamma distribution for scale-invariant squared error loss. *Canad. J. Statist.* **23**, pp. 245–256.
40. C. VAN EEDEN and J.V. ZIDEK (in preparation). Minimax estimation of a bounded scale parameter for scale-invariant squared error loss.
41. E. ZINZIUS (1981). Minimaxschätzer für den Mittelwert ϑ einer normalverteilten Zufallgröße mit bekannter Varianz bei vorgegebener oberer und unterer Schranke für ϑ. *Math. Operationsforsch. Statist.* **12**, pp. 551–557.