2

A statement of the problem, the notation and some definitions

Consider a probability space $(\mathcal{X}, \mathcal{A})$, a family of distributions

$$\mathcal{P}_o = \{P_{\theta,\lambda}, \theta = (\theta_1, \ldots, \theta_M), \lambda = (\lambda_1, \ldots, \lambda_{K-M}), (\theta, \lambda) \in \Omega_o \subset R^K\}$$

defined over it and a random vector $X \in R^n$ defined on it, where $M \geq 1$, $K - M \geq 0$ and Ω_o is closed and convex with a non-empty interior. Assume that $P_{\theta,\lambda}$ has a density, $p_{\theta,\lambda}$, with respect to a σ-finite measure ν. Then the problem considered in this monograph is the estimation, based on X, of θ when it is known that $(\theta, \lambda) \in \Omega$, where Ω is a known closed, convex proper subset of Ω_o with a non-empty interior. When $K - M \geq 1$, λ is a vector of nuisance parameters.

Let

$$\left.\begin{aligned} \Theta_o &= \{\theta \in R^M \mid (\theta, \lambda) \in \Omega_o \text{ for some } \lambda \in R^{K-M}\} \\ \Theta &= \{\theta \in R^M \mid (\theta, \lambda) \in \Omega \text{ for some } \lambda \in R^{K-M}\}. \end{aligned}\right\} \tag{2.1}$$

Then the set Θ is a known closed, convex subset of Θ_o with a non-empty interior.

For a definition of what, in this monograph, is considered to be an estimator of θ and to see its relationship to definitions used by other authors, first look at the case where

$$\text{the support of } P_{\theta,\lambda} \text{ is independent of } (\theta, \lambda) \text{ for } (\theta, \lambda) \in \Omega_o. \tag{2.2}$$

Then estimators δ of θ based on X satisfy

$$P_{\theta,\lambda}(\delta(X) \in \Theta) = 1 \text{ for all } (\theta, \lambda) \in \Omega. \tag{2.3}$$

This class of estimators is denoted by

$$\mathcal{D} = \{\delta \mid (2.3) \text{ is satisfied }\}. \tag{2.4}$$

As examples of this kind of model, let X_1 and X_2 be independent random variables with $X_1 \sim \mathcal{N}(\theta, 1)$ and $X_2 \sim \mathcal{N}(\lambda, 1)$. Suppose θ and λ are unknown, but it is known that $\theta \leq \lambda \leq 1$. Then $K = 2$, $M = 1$, $\Omega_o = R^2$, $\Omega = \{(\theta, \lambda) \mid \theta \leq \lambda \leq 1\}$, $\Theta_o = R^1$ and $\Theta = \{\theta \mid \theta \leq 1\}$. In this case the problem is the estimation of θ based on $X = (X_1, X_2)$ by an estimator $\delta(X)$ satisfying $P_{\theta,\lambda}(\delta(X) \leq 1) = 1$ for all $(\theta, \lambda) \in \Omega$. For the case where $X_i \sim^{ind} \mathcal{N}(\theta_i, 1)$, $i = 1, 2$, with θ_1 and θ_2 unknown and $\theta_1 \leq \theta_2$, $K = M = 2$ and $\theta = (\theta_1, \theta_2)$ is to be estimated based on $X = (X_1, X_2)$. Here $\Theta_o = \Omega_o = R^2$, $\Theta = \Omega = \{\theta \mid \theta_1 \leq \theta_2\}$ and estimators $\delta(X) = (\delta_1(X), \delta_2(X))$ satisfy $P_\theta(\delta_1(X) \leq \delta_2(X)) = 1$ for all $\theta \in \Theta$. As another example, let $X_i \sim^{ind} \mathcal{N}(\theta, \lambda_i)$, $i = 1, \ldots, k$ with all parameters unknown and $0 < \lambda_1 \leq \ldots \leq \lambda_k$. Then θ is to be estimated based on $X = (X_1, \ldots, X_k)$. Here $K = k+1$, $M = 1$, $\Omega_o = \{(\theta, \lambda_1, \ldots, \lambda_k) \mid -\infty < \theta < \infty, \lambda_i > 0, i = 1, \ldots, k\}$, $\Omega = \{(\theta, \lambda_1, \ldots, \lambda_k) \mid -\infty < \theta < \infty, 0 < \lambda_1 \leq \ldots \leq \lambda_k\}$, and $\Theta_o = \Theta = R^1$.

Not every author on the subject of estimation in restricted parameter spaces restricts his estimators of θ to those satisfying (2.3). Some authors ask, for some or all of their estimators, only that they satisfy

$$P_{\theta,\lambda}(\delta(X) \in \Theta_o) = 1 \text{ for all } (\theta, \lambda) \in \Omega. \tag{2.5}$$

Others do not say what they consider to be an estimator, but their definition can sometimes be obtained from the properties they prove their estimators to have. A summary of opinions on whether estimators should satisfy (2.3) can be found in Blyth (1993). Here I only quote Hoeffding (1983) on the subject of restricting estimators of θ to those in $\mathcal{D}$. He calls such estimators "range-preserving" and says "The property of being range-preserving is an essential property of an estimator, a sine qua non. Other properties, such as unbiasedness, may be desirable in some situations, but an unbiased estimator that is not range-preserving should be ruled out as an estimator." – a statement with which I agree.

Let $L(d, \theta)$ be the loss incurred when $d \in \Theta$ is used to estimate θ and $\theta \in \Theta$ is the true value of the parameter to be estimated. It is assumed that L is of the form

$$L(d, \theta) = \sum_{i=1}^{M} L_i(d_i, \theta_i), \tag{2.6}$$

where, for each $i = 1, \ldots, M$, all $y = (y_1, \ldots, y_M) \in \Theta$ and all $(\theta_1, \ldots, \theta_M) \in \Theta$,

$$\left.\begin{aligned} &i)\ L_i(y_i, \theta_i) \text{ is bowl-shaped in } \theta_i,\\ &ii)\ L_i(y_i, \theta_i) \geq 0 \text{ and } L_i(\theta_i, \theta_i) = 0,\\ &iii)\ L_i(y_i, \theta_i) \text{ is convex in } y_i. \end{aligned}\right\} \tag{2.7}$$

These properties of the loss function, together with the convexity of Θ, imply that the class of non-randomized estimators is essentially complete in the class of all estimators with respect to Ω in the sense that, for every randomized estimator δ, there exists a non-randomized one δ' with $\mathcal{E}_{\theta,\lambda} L(\delta'(X), \theta) \leq L(\delta(X), \theta)$ for all $(\theta, \lambda) \in \Omega$. So one can restrict oneself to non-randomized estimators.

Examples of loss functions of the form (2.6) with the properties (2.7) are

1) the class of weighted p^{th}-power loss functions where

$$L(d, \theta) = \sum_{i=1}^{M} |d_i - \theta_i|^p w_i(\theta).$$

Here $p \geq 1$ and the $w_i(\theta)$ are known functions of θ which are, for each $i = 1, \ldots, M$, strictly positive on Θ. Special cases of this loss function are (i) squared-error loss with $p = 2$ and $w_i(\theta) = 1$ for all $i = 1, \ldots, M$ and (ii) scale-invariant squared-error loss with $p = 2$ and $w_i(\theta) = 1/\theta_i^2$, which can be used when $\theta_i > 0$ for all $\theta \in \Theta$, as is, e.g., the case in scale-parameter estimation problems;

2) the class of linex loss functions where

$$L(d, \theta) = \sum_{i=1}^{M} \left(e^{w_i(\theta)(d_i - \theta_i)} - w_i(\theta)(d_i - \theta_i) - 1 \right).$$

Here the $w_i(\theta)$ are known functions of θ with, for each $i = 1, \ldots, M$, $w_i(\theta) \neq 0$ for all $\theta \in \Theta$.

In problems with $M \geq 2$ quadratic loss is sometimes used. It generalizes squared-error loss and is given by

$$L(d, \theta) = (d - \theta)' A (d - \theta), \tag{2.8}$$

where A is a known $M \times M$ positive definite matrix. For instance, when $X \sim \mathcal{N}_M(\theta, \Sigma)$, Σ known and positive definite, with the vector θ to be estimated, taking $A = \Sigma^{-1}$ is equivalent to estimating the vector $\Sigma^{-1/2}\theta$ with squared-error loss based on $Y = \Sigma^{-1/2} X$.

The risk function of an estimator δ of θ is, for $(\theta, \lambda) \in \Omega$, given by $R(\delta, (\theta, \lambda)) = \mathcal{E}_{\theta,\lambda} L(\delta(X), \theta)$ and estimators are compared by comparing their risk functions. An estimator δ is called inadmissible in a class $\mathcal{C}$ of estimators for estimating θ if there exists an estimator $\delta' \in \mathcal{C}$ dominating it on Ω, i.e., if there exists an estimator $\delta' \in \mathcal{C}$ with

$$R(\delta', (\theta, \lambda)) \leq R(\delta, (\theta, \lambda)) \text{ for all } (\theta, \lambda) \in \Omega \text{ and}$$

$$R(\delta', (\theta, \lambda)) < R(\delta, (\theta, \lambda)) \text{ for some } (\theta, \lambda) \in \Omega$$

and an estimator δ is admissible when it is not inadmissible. Further, an estimator δ of θ is called minimax in a class $\mathcal{C}$ of estimators of θ if it minimizes, among estimators $\delta' \in \mathcal{C}$, $\sup_{(\theta,\lambda)\in\Omega} R(\delta', (\theta, \lambda))$.

In the literature on estimation in restricted parameter spaces two definitions of admissibility and minimaxity are used. In each of the definitions the risk functions of estimators of θ are compared on Ω. However, in one definition the estimators under consideration (i.e., the above class $\mathcal{C}$) are those in $\mathcal{D}$, while in the other definition the estimators are those in $\mathcal{D}_o = \{\delta \mid (2.5) \text{ is satisfied}\}$. Or, to say this another way, by the first definition an estimator δ ($\in \mathcal{D}$) is inadmissible when there exists an estimator $\delta' \in \mathcal{D}$ which dominates δ on Ω. And an estimator δ ($\in \mathcal{D}$) is minimax when it minimizes, among estimators in $\mathcal{D}$, $\sup\{R(\delta, (\theta, \lambda)) \mid (\theta, \lambda) \in \Omega\}$. By the second definition, an estimator δ ($\in \mathcal{D}_o$) is inadmissible if there exists an estimator $\delta' \in \mathcal{D}_o$ which dominates it on Ω. And an estimator δ ($\in \mathcal{D}_o$) is minimax if it minimizes, among the estimators in $\mathcal{D}_o$, $\sup\{R(\delta, (\theta, \lambda)) \mid (\theta, \lambda) \in \Omega\}$. It is hereby assumed, for the second pair of definitions, that the loss function (2.6)

$$\left.\begin{array}{l} i) \text{ is defined for } \theta \in \Theta \text{ and } d \in \Theta_o \\ ii) \text{ satisfies (2.7) with } \theta \in \Theta \text{ and } (y_1, \ldots, y_M) \in \Theta_o. \end{array}\right\} \tag{2.9}$$

These two notions of admissibility and minimaxity will be called, respectively, $(\mathcal{D}, \Omega)$- and $(\mathcal{D}_o, \Omega)$-admissibility and minimaxity and the corresponding estimation problems will be called, respectively, the $(\mathcal{D}, \Omega)$- and the $(\mathcal{D}_o, \Omega)$-problems. In this monograph estimators satisfy (unless specifically stated otherwise) (2.3) and admissibility and minimaxity mean $(\mathcal{D}, \Omega)$-admissibility and minimaxity.

The following relationships exist between $(\mathcal{D}, \Omega)$- and $(\mathcal{D}_o, \Omega)$-admissibility and minimaxity:

$$\delta \in \mathcal{D}, \delta \text{ is } (\mathcal{D}_o, \Omega)\text{-admissible} \implies \delta \text{ is } (\mathcal{D}, \Omega)\text{-admissible.} \tag{2.10}$$

Further,

$$\delta \in \mathcal{D}, \delta \text{ is } (\mathcal{D}_o, \Omega)\text{-minimax} \implies \begin{cases} \delta \text{ is } (\mathcal{D}, \Omega)\text{-minimax,} \\ M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega), \end{cases} \tag{2.11}$$

where $M(\mathcal{D}, \Omega)$ and $M(\mathcal{D}_o, \Omega)$ are the minimax values for the classes $\mathcal{D}$ and $\mathcal{D}_o$ and the parameter space Ω.

Now note that, for weighted squared-error loss, the class of estimators $\mathcal{D}$ is essentially complete in $\mathcal{D}_o$ with respect to the parameter space Ω in the sense that, for every $\delta \in \mathcal{D}_o$, $\delta \notin \mathcal{D}$ there exists a $\delta' \in \mathcal{D}$ dominating it on Ω. This dominator is obtained by minimizing, for each $x \in \mathcal{X}$, $L(\delta(x), \theta)$ in θ for

$\theta \in \Theta$. This essential completeness also holds for a rectangular Θ when the loss function is given by (2.6) and satisfies (2.9). Further, one can dominate δ by using what Stahlecker, Knautz and Trenkler (1996) call the "minimax adjustment technique". Their dominator – δ', say – is obtained by minimizing, for each $x \in \mathcal{X}$,

$$H(d) = \sup_{\theta \in \Theta} (L(d, \theta) - L(\delta(x), \theta))$$

for $d \in \Theta$. When $\delta(x) \in \Theta$, $H(d) \geq 0$ because

i) $H(\delta(x)) = 0$, so $\inf_{d \in \Theta} H(d) \leq 0$;
ii) $\inf_{d \in \Theta} H(d) < 0$ contradicts the fact that, for each $d \in \Theta$,

$$\begin{aligned} H(d) &= \sup_{\theta \in \Theta} (L(d, \theta) - L(\delta(x), \theta)) \\ &\geq L(d, \delta(x)) - L(\delta(x), \delta(x)) = L(d, \delta(x)) \geq 0. \end{aligned}$$

So, when $\delta(x) \in \Theta$, $d = \delta(x)$ is a minimizer of $H(d)$. When $\delta(x)$ is not in Θ, assume that a minimizer exists. Then we have, for all $\theta \in \Theta$ and all $x \in \mathcal{X}$,

$$\left. \begin{array}{l} L(\delta'(x), \theta) - L(\delta(x), \theta) = \\ \inf_{d \in \Theta} \sup_{\theta \in \Theta} (L(d, \theta) - L(\delta(x), \theta)) = \\ \sup_{\theta \in \Theta} \inf_{d \in \Theta} (L(d, \theta) - L(\delta(x), \theta)) = \\ \sup_{\theta \in \Theta} (-L(\delta(x), \theta)) = -\inf_{\theta \in \Theta} L(\delta(x), \theta) \leq 0, \end{array} \right\} \quad (2.12)$$

where it is assumed that inf and sup can be interchanged.

Essential completeness of $\mathcal{D}$ in $\mathcal{D}_o$ with respect to Ω, together with (2.10) gives

$$\delta \text{ is } (\mathcal{D}, \Omega)\text{-admissible} \iff \delta \in \mathcal{D}, \delta \text{ is } (\mathcal{D}_o, \Omega)\text{-admissible}. \quad (2.13)$$

Further, using (2.11) and the essential completeness of $\mathcal{D}$ in $\mathcal{D}_o$ with respect to Ω, one obtains

$$\left. \begin{array}{c} \delta \text{ is } (\mathcal{D}, \Omega)\text{-minimax} \Leftrightarrow \delta \in \mathcal{D}, \delta \text{ is } (\mathcal{D}_o, \Omega)\text{-minimax}. \\ M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega). \end{array} \right\} \quad (2.14)$$

From (2.13) and (2.14) it is seen that studying the $(\mathcal{D}_o, \Omega)$-problem can be very helpful for finding admissibility and minimaxity results for the $(\mathcal{D}, \Omega)$-problem.

Another problem whose admissibility and minimaxity results can be helpful for our problem is the "unrestricted problem" where estimators of θ are restricted to Θ_o and compared on Ω_o. Then we have (still assuming that (2.2) holds)

$$\delta \in \mathcal{D}_o \Longrightarrow P_{\theta,\lambda}(\delta(X) \in \Theta_o) = 1 \text{ for all } (\theta, \lambda) \in \Omega_o,$$

so that this estimation problem can, and will, be called the $(\mathcal{D}_o, \Omega_o)$-problem. Obviously,

$$\delta \in \mathcal{D}_o, \delta \text{ is } (\mathcal{D}_o, \Omega)\text{-admisible} \Longrightarrow \delta \in \mathcal{D}_o, \delta \text{ is } (\mathcal{D}_o, \Omega_o)\text{-admissible}. \tag{2.15}$$

Also, because $\Omega \subset \Omega_o$,

$$M(\mathcal{D}_o, \Omega) \leq M(\mathcal{D}_o, \Omega_o) \tag{2.16}$$

which, together with (2.14), gives

$$M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega) \leq M(\mathcal{D}_o, \Omega_o). \tag{2.17}$$

One can now ask the question: when does

$$M(\mathcal{D}_o, \Omega) = M(\mathcal{D}_o, \Omega_o) \tag{2.18}$$

or, equivalently,

$$M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega_o) \tag{2.19}$$

or, equivalently

$$M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega) = M(\mathcal{D}_o, \Omega_o) \tag{2.20}$$

hold? Or – are there cases where restricting the parameter space does not reduce the minimax value of the problem?

Examples where (2.2) and (2.20) hold can be found in Chapter 4, sections 4.2, 4.3 and 4.4. Those in the sections 4.3 and 4.4 are examples where either all the θ_i are lower-bounded or $\Theta = \{\theta \mid \theta_1 \leq \ldots, \leq \theta_k\}$. In the example in Section 4.2, $X \sim \text{Bin}\,(n, \theta)$ with $\theta \in [m, 1-m]$ for a small known $m \in (0, 1/2)$.

An example where (2.2) is satisfied but (2.20) does not hold is the estimation of a bounded normal mean with $\Theta_o = (-\infty, \infty)$, squared-error loss and known variance. When $\Theta = [-m, m]$ for some positive known m, Casella and Strawderman (1981) give the values of $M(\mathcal{D}, \Omega)$ for several values of m. For example, for $m = .1, .5, 1$ and a normal distribution with variance 1, the minimax risks are, respectively $.010, .199, .450$, while, of course, $M(\mathcal{D}_o, \Omega_o) = 1$, showing that restricting the parameter space to a compact set can give very substantial reductions in the minimax value of the problem. These results are discussed in Chapter 4, Section 4.2 together with other cases where the three minimax values are not equal.

As already noted, the above relationships between admissibility and minimaxity results for the $(\mathcal{D}, \Omega)$- and $(\mathcal{D}_o, \Omega)$-problems show that solving a $(\mathcal{D}_o, \Omega)$-problem can be very helpful toward finding a solution to the corresponding $(\mathcal{D}, \Omega)$-problem. But authors who publish results on a $(\mathcal{D}_o, \Omega)$-problem are not always clear about why they do so. Is it as a help for solving the corresponding $(\mathcal{D}, \Omega)$-problem, or do they consider statistics not satisfying (2.3) to be estimators and are not really interested in the corresponding $(\mathcal{D}, \Omega)$-problem? In this monograph some papers are included which look at $(\mathcal{D}_o, \Omega)$-problems. Their results are clearly identified as such, but their relationship to the corresponding $(\mathcal{D}, \Omega)$-problems is not always commented on.

Remark 2.1. Note that, when $\delta \in \mathcal{D}_o$ is $(\mathcal{D}_o, \Omega_o)$-minimax and $\delta' \in \mathcal{D}$ dominates δ on Ω, one cannot conclude that δ' is $(\mathcal{D}, \Omega)$-minimax. But Dykstra (1990) seems, in his Example 3, to draw this conclusion.

Remark 2.2. The above definition of the minimax adjustment technique is not the one used by Stahlecker, Knautz and Trenkler (1996). They minimize, for each $x \in \mathcal{X}$, $H(d)$ for $d \in R^k$. Such a minimizer is not necessarily constrained to Θ. But assuming they meant to minimze over Θ, their reasoning is incorrect.

In most papers on restricted-parameter-space estimation the models considered satisfy (2.2), but several models where this condition is not satisfied are rather extensively studied. Three examples are the k-sample problems where $X_{i,1}, \ldots, X_{i,n_i}$, $i = 1, \ldots, k$, are independent and $X_{i,j}$, $j = 1, \ldots, n_i$ have either a $\mathcal{U}(0, \theta_i)$ distribution or a $\mathcal{U}(\theta_i - 1, \theta_i + 1)$ distribution or an exponential distribution on (θ_i, ∞). Suppose, in the first uniform case, that $\theta = (\theta_1, \ldots, \theta_k)$ is to be estimated when $\theta \in \Theta$, where Θ is a closed convex subset of R^k_+ with a non-empty interior. Then $M = k = K$, $\Omega_o = \Theta_o = R^k_+$ and $\Omega = \Theta$. Given that we know for sure, i.e., with $P_\theta = 1$ for all $\theta \in R^k_+$, that $Y_i = \max_{1 \le j \le n_i} X_{i,j} \le \theta_i$, $i = 1, \ldots, k$, estimators δ of θ "should", in addition to being restricted to Θ, satisfy the "extra" restriction that $\delta_i(Y) \ge Y_i$, $i = 1, \ldots, k$, where $Y = (Y_1, \ldots, Y_k)$. To say it more precisely, δ should satisfy

$$P_\theta(\delta(Y) \in \Theta_Y) = 1 \text{ for all } \theta \in \Theta, \tag{2.21}$$

where

$$\Theta_Y = \{\theta \in \Theta \mid \theta_i \ge Y_i, i = 1, \ldots, k\}.$$

Let $\mathcal{D}'$ be the class of estimators satisfying (2.21) and let the $(\mathcal{D}', \Theta)$-problem be the problem where estimators are in $\mathcal{D}'$ and are compared on Θ. The unrestricted problem in this case is the problem where estimators satisfy

$$P_\theta(\delta(Y) \in \Theta_{o,Y}) = 1 \text{ for all } \theta \in R^k_+, \tag{2.22}$$

where

$$\Theta_{o,Y} = \{\theta \in R^k_+ \mid \theta_i \ge Y_i, i = 1, \ldots, k\}$$

and estimators are compared on R_+^k. Call this problem the $(\mathcal{D}'_o, R_+^k)$-problem. And then there is the problem studied by those who do not insist that their estimators are restricted to Θ, i.e., the $(\mathcal{D}'_o, \Theta)$-problem where estimators satisfy (2.22) and are compared on Θ.

From the above definitions it follows that $\Theta_Y \subset \Theta_{o,Y}$ with $P_\theta = 1$ for all $\theta \in \Theta$ and that Θ_Y and $\Theta_{o,Y}$ are both closed and convex with a non-empty interior. So, $\mathcal{D}' \subset \mathcal{D}'_o$ and, under the same conditions on the loss function as before, $\mathcal{D}'$ is essentially complete in $\mathcal{D}'_o$ with respect to Θ. This shows that (2.13) and (2.14) hold with $\mathcal{D}$ (resp. $\mathcal{D}_o$) replaced by $\mathcal{D}'$ (resp. $\mathcal{D}'_o$). With these same replacements, (2.15) holds for the $(\mathcal{D}'_o, R_+^k)$- and the $(\mathcal{D}'_o, \Theta)$-problems so that (see (2.17)),

$$M(\mathcal{D}', \Theta) = M(\mathcal{D}'_o, \Theta) \leq M(\mathcal{D}'_o, R_+^k). \tag{2.23}$$

An example where these three minimax values are equal is given in Chapter 4, Section 4.3.

Similar remarks and results hold for the other uniform case and for the exponential case, as well as for cases with nuissance parameters.

In order to simplify the notation, $(\mathcal{D}, \Omega)$, $(\mathcal{D}_o, \Omega)$ and $(\mathcal{D}_o, \Omega_o)$ are used for the three problems (with the Ω's replaced by Θ's in case there are no nuissance parameters), whether (2.2) is satisfied or not: i.e., the primes are left off for cases like the uniform and exponentail ones above. And "δ satisfies (2.3)" stands for "δ satisfies (2.3) or (2.21)", as the case may be.

Quite a number of papers on such uniform and exponential models are discussed in this monograph. In most of them the "extra" restriction is taken into account but, as will be seen in Chapter 5, Sections 5.2 and 5.3, in two cases authors propose and study estimators which do not satisfy it.

Two more remarks about admissibility and minimaxity for the three problems: (i) if δ is $(\mathcal{D}, \Omega)$-inadmissible as well as $(\mathcal{D}, \Omega)$-minimax, then every $\delta' \in \mathcal{D}$ which dominates δ on Ω is also $(\mathcal{D}, \Omega)$-minimax. This also holds with $(\mathcal{D}, \Omega)$ replaced by $(\mathcal{D}_o, \Omega)$ as well with $(\mathcal{D}, \Omega)$ replaced by $(\mathcal{D}_o, \Omega_o)$; (ii) if δ is $(\mathcal{D}_o, \Omega)$-minimax and $\mathcal{D}$ is essentially complete in $\mathcal{D}_o$ with respect to Ω then there exists a $\delta' \in \mathcal{D}$ which is $(\mathcal{D}, \Omega)$-minimax, because the essential completeness implies that $M(\mathcal{D}, \Omega) = M(\mathcal{D}_o, \Omega)$.

Universal domination is another criterion for comparing estimators. It was introduced by Hwang (1985) for the case where $\Theta = \Omega = R^k$ and by this criterion an estimator δ' universally dominates an estimator δ on the parameter space Θ with respect to a class $\mathcal{C}$ of loss functions L if

$$\mathcal{E}_\theta L(\delta'(X), \theta) \leq \mathcal{E}_\theta L(\delta(X), \theta) \quad \text{for all } \theta \in \Theta \text{ and all } L \in \mathcal{C}$$

and, for a particular loss function $\in \mathcal{C}$, the risk functions are not identical. An estimator δ is called universally admissible if no such δ' exists.

Hwang (1985) takes the class $\mathcal{C}$ to be the class of all nondecreasing functions of the generalized Euclidean distance $|d-\theta|_D = ((d-\theta)'D(d-\theta))^{1/2}$ where D is a given non-negative definite matrix. This implies, as he shows, that δ' universally dominates δ if and only if δ' stochastically dominates δ, i.e. if and only if

$$P_\theta(|\delta'(X)-\theta|_D \geq c) \leq P_\theta(|\delta(X)-\theta|_D \geq c) \qquad \text{for all } c>0 \text{ and all } \theta \in \Theta$$

$$P_\theta(|\delta'(X)-\theta|_D \geq c) < P_\theta(|\delta(X)-\theta|_D \geq c) \qquad \text{for some } (c,\theta), c>0, \theta \in \Theta.$$

He further shows that, if δ is admissible with respect to a particular loss function L_o which is a strictly increasing function of $|d-\theta|_D$ and the risk function of δ for this loss function is finite for all $\theta \in \Theta$, then δ is universally admissible for this D and Θ. Equivalently, if δ is universally inadmissible with respect to a D and Θ, then δ is inadmissible under any strictly increasing loss $L_o(|d-\theta|_D)$ with a risk which is finite for all $\theta \in \Theta$.

Still another criterion for comparing estimators is Pitman closeness (also called Pitman nearness). For two estimators δ_1 and δ_2 of $\theta \in \Theta \subset R^1$, Pitman (1937) defines, for cases where $K=M$, their closeness by

$$P_\theta(|\delta_1(X)-\theta| < |\delta_2(X)-\theta|) \qquad \theta \in \Theta. \tag{2.24}$$

Then, assuming that $P_\theta(|\delta_1(X)-\theta| = |\delta_2(X)-\theta|) = 0$ for all $\theta \in \Theta$, δ_1 is Pitman-closer to θ than δ_2 when (2.24) is $\geq 1/2$. Pitman (1937) notes that Pitman closeness comparisons are not necessarily transitive. For three estimators δ_i, $i=1,2,3$, one can have δ_1 Pitman-closer to θ than δ_2, δ_2 Pitman-closer to θ than δ_3 and δ_3 Pitman-closer to θ than δ_1. It is also well-known that Pitman-closeness comparisons do not necessarily agree with risk-function comparisons. One can have, e.g., δ_1 Pitman-closer to θ than δ_2 while δ_2 dominates δ_1 for squared-error loss. In Chapter 5, Section 5.3, several Pitman-closeness comparisons in restricted parameter spaces are presented and compared with risk-function comparisons. Much more on Pitman closeness, in particular on its generalization to $k \geq 2$, can be found in Keating, Mason and Sen (1993).

One of the various estimators discussed in this monograph is the so-called Pitman estimator. The name comes from Pitman (1939). He proposes and studies Bayes estimators with respect to a uniform prior. His parameters are either location parameters $\theta \in \Theta_o = (-\infty, \infty)$, for which he uses squared-error loss and a uniform prior on Θ, or scale parameters $\theta \in \Theta_o = (0, \infty)$, for which he uses scale-invariant-squared-error loss and a uniform prior for $\log\theta$ on $(-\infty,\infty)$. But the name "Pitman estimator" is now used by many authors, and is used in this monograph, for any Bayes estimator with respect to a uniform prior for θ or for a function $h(\theta)$ for $\theta \in \Theta$ or Θ_o. Some of the properties of the original Pitman estimators are summarized in Chapter 4, Section 4.1.

In many restricted-parameter-space estimation problems considered in the literature the problem does not contain any nuisance parameters, the problem is a k-sample problem with independent samples from distributions $F_i(x, \theta_i)$, $i = 1, \ldots, k$ and Θ $(= \Omega)$ is determined by inequalities among the components θ_i of θ. The most common examples are the simple-order restriction where $\Theta = \{\theta \mid \theta_1 \leq \ldots \leq \theta_k\}$, the simple-tree-order restriction with $\Theta = \{\theta \mid \theta_1 \leq \theta_i, i = 2, \ldots, k\}$, the umbrella-order restriction with $\Theta = \{\theta \mid \text{for some } i_o, 1 < i_o < k, \theta_i \leq \theta_{i_o} \text{ for all } i \neq i_o\}$ and the loop-order restriction with, for $k = 4$ e.g., $\Theta = \{\theta \mid \theta_1 \leq \theta_2 \leq \theta_4, \theta_1 \leq \theta_3 \leq \theta_4\}$. The simple-tree-order restriction is a special case of the rooted-tree-order restriction where each θ_i, except one of them (θ_1, say, the root), has exactly one immediate predecessor and the root has none. Here, θ_j is an immediate predecessor of θ_i $(i \neq j)$ when $\theta \in \Theta$ implies $\theta_j \leq \theta_i$ but there does not exist an l, $l \neq i$, $l \neq j$, with $\theta_j \leq \theta_l \leq \theta_i$. So, the simple-tree order is a tree order where all θ_i have the root as their unique immediate predecessor. Another Θ for which results have been obtained is the upper-star-shaped restriction, also called the increasing-in-weighted-average restriction, where

$$\Theta = \{\theta \mid \theta_1 \leq \bar{\theta}_2 \leq \ldots \leq \bar{\theta}_k\},$$

with $\bar{\theta}_i = \sum_{j=1}^{i} w_j \theta_j / \sum_{j=1}^{i} w_j$ for given positive weights w_i. Note that this Θ is equivalent to

$$\Theta = \{\theta \mid \bar{\theta}_i \leq \theta_{i+1}, i = 1, \ldots, k-1\}.$$

Finally, when $\theta_1, \ldots, \theta_k$ are order-restricted, θ_i is a node when, for each $j \neq i$, $\theta_j \leq \theta_i$ or $\theta_j \geq \theta_i$. For instance, when $k = 5$ and

$$\Theta = \{\theta \mid \theta_j \leq \theta_3, j = 1, 2, \theta_l \geq \theta_3, l = 4, 5\},$$

θ_3 is the only node and when

$$\Theta = \{\theta \mid \theta_j \leq \theta_3, j = 1, 2, \theta_3 \leq \theta_4 \leq \theta_5\}$$

then θ_3, θ_4 and θ_5 are nodes, but θ_1 and θ_2 are not. Nodes are important for some estimation problems presented in Chapter 5.