On the Minimax Point Estimations

Miyasawa, Koichi
Akita University

https://doi.org/10.5109/12958

出版情報：統計数理研究. 5 (3/4), pp.1-17, 1953-06. Research Association of Statistical Sciences

ON THE MINIMAX POINT ESTIMATIONS

By Kôichi MIYASAWA

§ 1. Introduction

Let $X = \{X_i\}$ $(i = 1, 2, \cdots)$ be a denumerable sequence of random variables X_i, x_i be the observed value of X_i, and $x = \{x_i\}$. Let $F(x)$ be the distribution function of X, and Ω be the given class of possible distribution functions of X. We assume that the true distribution function $F(x)$ of X is not known, but we merely know that the distribution function $F(x)$ of X is an element of Ω. Let ξ_F be a probability measure on Ω which gives probability 1 to the specified element F of Ω, and Ω^* be the set of all such probability measures ξ_F. We write a probability measure ξ on Ω, which gives probabilities α^i, $i = 1, 2, \cdots, n$, $\alpha^i \geq 0$, $\sum_{i=1}^{n} \alpha^i = 1$, to some finite number of specified elements F^i $(i = 1, 2, \cdots, n)$, as $\xi = \sum_{i=1}^{n} \alpha^i \xi_{F^i}$, and the set of all these probability measures ξ as P. Let d be a decision function $(d.f.)$ and D be the set of all possible $d.f.$'s.

Let $r(F, d)$ be the risk when F is the true distribution function of X and d is the adopted $d.f.$, and $r(F, d)$ be the given non-negative bounded function of F and d. We define the metric between two elements d_1 and d_2 of D as follows

$$(1) \qquad \rho(d_1, d_2) = \sup_F |r(F, d_1) - r(F, d_2)|.$$

Let B be the Borel field of subsets of D, generated by open subsets (with respect to the metric (1)) of D, and η be a probability measure on D.

Let R be the set of all these probability measures η. An element d of D is a non-randomized $d.f.$ and an element η of R is a randomized $d.f.$

We assume that, for any fixed F, $r(F, d)$ is a measurable B function of d. We write

$$(2) \qquad r(\xi, \eta) = \iint r(F, d)\, d\xi\, d\eta.$$

If it holds

$$(3) \qquad \inf_\eta \sup_\xi r(\xi, \eta) = \sup_\xi \inf_\eta r(\xi, \eta) \quad (= \lambda_0, \text{ say})$$

then we say that the decision problem is strictly determined.

Definition 1. If the decision problem is strictly determined, and if η_0 is a $d.f.$ such that

---

(1) Communicated at the Autumn-Meeting of Japanese Math. Soc., at Kyôto Univ., November 4, 1952.

$$(4) \qquad \lambda_0 = \sup_{\xi} r(\xi, \eta_0) = \min_{\eta} \sup_{\xi} r(\xi, \eta),$$

then we define η_0 as a minimax solution of the decision problem.

§ 2. The general Theory*.

We define the operator T on the space R by the following

$$(5) \qquad T\eta = \int r(F, d)\, d\eta = r(F, \eta) \qquad \text{for all } \eta \in R.$$

We consider ξ of P as a functional on TR and write

$$(6) \qquad (\xi, T\eta) = \int r(F, \eta)\, d\xi = r(\xi, \eta).$$

For any positive integer n, let

$$(7) \quad S_n = \{\alpha\, ;\ \alpha = (\alpha^1, \cdots, \alpha^n),\ \alpha^i \geq 0,\ i = 1, \cdots, n,\ \textstyle\sum_{i=1}^{n} \alpha^i = 1\}$$

Lemma. 1. Let $\xi_{F_1}, \cdots, \xi_{F_n}$ be any fixed finite number of elements of Ω^* and $C = [\eta_1, \cdots, \eta_k]$ be a convex set spanned by any fixed finite number of elements $\eta_1, \cdots, \eta_k$ of R. If for any $\eta \in C$, there exists a ξ_{F_i} with

$$(\xi_{F_i}, T\eta) \geq \lambda$$

where λ is a real number, then there exists $\alpha_0 \in S_n$ such that

$$(\textstyle\sum \alpha_0^i \xi_{F_i}, T\eta) \geq \lambda \qquad \text{for all } \eta \in C.$$

Proof. We consider the matrix

$$a_{ij} = (\xi_{F_i}, T\eta_j), \quad i = 1, 2, \cdots, n; \quad j = 1, \cdots, k.$$

Then the condition of Lemma 1 states that for any $\beta \in S_k$, there exists an i such that

$$(8) \qquad (\xi_{F_i}, T \textstyle\sum_{j=1}^{k} \beta^j \eta_j) = \displaystyle\sum_{j=1}^{k} a_{ij} \beta^j \geq \lambda.$$

Accordingly we have

$$(9) \qquad \min_{\beta \in S_k} \max_{\alpha \in S_n} \sum_{i,j} \alpha^i a_{ij} \beta^j \geq \lambda.$$

On the other hand, since the finite game, generated by the matrix $\|a_{ij}\|$, is strictly determined, we have

$$(10) \qquad \min_{\beta} \max_{\alpha} \sum_{i,j} \alpha^i a_{ij} \beta^j = \max_{\alpha} \min_{\beta} \sum_{i,j} \alpha^i a_{ij} \beta^j.$$

Therefore, from (9), (10), we have

$$(11) \qquad \max_{\alpha \in S_n} \min_{\beta \in S_k} \sum_{i,j} \alpha^i a_{ij} \beta_j \geq \lambda.$$

---

* To prove Theorem 1 we shall apply the method used in [1].

Accordingly there exists $\alpha_0 \in S_n$ such that

$$\sum_{i,j} \alpha_0^i a_{ij} \beta^j = (\sum \alpha_0^i \xi_{F_i},\ T \sum \beta^j \eta_j) \geq \lambda$$

for all $\beta \in S_k$. This implies the conclusion of Lemma 1.

Lemma 2. Let $\xi_{F_1}, \cdots, \xi_{F_n}$ be any fixed finite number of elements of Ω^*. If for every $\eta \in R$, there exists a ξ_{F_i} with

$$(\xi_{F_i},\ T\eta) \geq \lambda$$

then there exists $\alpha_0 \in S_n$ such that

$$(\sum_i \alpha_0^i \xi_{F_i},\ T\eta) \geq \lambda \qquad \text{for all } \eta \in R.$$

Proof. Since, using Lemma 1, this can be shown as in [1], we shall omit the proof here.

Definition 2. Let

$$\lambda(\eta) = \sup_{\xi_F} (\xi_F, T\eta). \tag{12}$$

Let L be the set of $\lambda(\eta)$ for all $\eta \in R$, and

$$\lambda_0 = \inf_\eta \lambda(\eta) = \inf_\eta \sup_F r(F, \eta). \tag{13}$$

Definition 3. We say that T is completely continuous (*c.c.*) if for any sequence $\{\eta_n\}$ of elements η_n of R, there exists a subsequence $\{\eta_{n_k}\}$ and $\eta_0 \in R$ such that

$$\lim_{k\to\infty} (\xi_F, T\eta_{n_k}) = (\xi_F, T\eta_0) \qquad \text{uniformly in } \xi_F. \tag{14}$$

Lemma 3. If T is *c.c.*, then there exists $\eta_0 \in R$ such that

$$\lambda_0 = \lambda(\eta_0) = \min_\eta \sup_F r(F, \eta). \tag{15}$$

Proof. Let $\{\lambda_n\}$ be a decreasing sequence of elements λ_n of L with $\lim_{n\to\infty} \lambda_n = \lambda_0$. Then by the definition of λ_n, there exists $\eta_n \in R$ with

$$\lambda_n = \sup_{\xi_F} (\xi_F, T\eta_n). \tag{16}$$

Since T is *c.c.*, there exists a subsequence $\{\eta_{n_k}\}$ and $\eta_0 \in R$ such that

$$\lim_{k\to\infty} (\xi_F, T\eta_{n_k}) = (\xi_F, T\eta_0) \qquad \text{uniformly in } \xi_F. \tag{17}$$

Accordingly, from (16) and (17), we have

$$\sup_{\xi_F} (\xi_F, T\eta_0) = \lambda_0,$$

that is $\lambda_0 = \lambda(\eta_0)$.

Lemma. 4. If, for evry $\xi \in P$, there exists $\eta_\xi \in R$ such that

$$(\xi, T\eta_\xi) < \lambda, \tag{18}$$

then for every fixed finite number of elements $\xi_1, \cdots, \xi_n$ of P, there exists $\eta \in R$ such that

$$(\xi_i, T\eta) < \lambda \qquad \text{for } i = 1, \cdots, n. \tag{19}$$

Proof. At first, we shall prove Lemma 4 for every fixed finite number of elements $\xi_{F_1}, \cdots, \xi_{F_m}$ of Ω^*, that is, there exists $\eta \in R$ such that

$$(\xi_{F_i}, T\eta) < \lambda \qquad \text{for } i = 1, \cdots, m. \tag{20}$$

If (20) does not holds, then there exist certain finite number of elements $\xi_{F_1}, \cdots, \xi_{F_k}$ such that, for every $\eta \in R$, there exists an ξ_{F_i} with

$$(\xi_{F_i}, T\eta) \geq \lambda . \tag{21}$$

Then by Lemma 2, there exists $\alpha_0 \in S_k$ such that

$$(\sum \alpha_0^i \xi_{F_i}, T\eta) \geq \lambda \qquad \text{for all } \eta \in R. \tag{22}$$

Now, since $\sum \alpha_0^i \xi_{F_i} \in P$, (22) contradicts the condition of Lemma 4. Therefore (20) is proved. Now let $\xi_1, \cdots, \xi_n$ be elements of P. Then we can write

$$\xi_i = \sum_{\nu=1}^{m_i} \alpha_\nu^i \xi_{F_{i\nu}}, \quad i = 1, 2, \cdots, n; \quad \alpha_i \in S_{m_i} .$$

For these $\xi_{F_{i\nu}}$, $i = 1, \cdots, n$; $\nu = 1, \cdots, m_i$, from (20), there exists $\eta \in R$ such that

$$(\xi_{F_{i\nu}}, T\eta) < \lambda \quad \text{for} \quad i = 1, \cdots, n; \quad \nu = 1, \cdots, m_i .$$

Then we have

$$(\xi_i, T\eta) = \sum \alpha_i^\nu (\xi_{F_{i\nu}}, T\eta) < \lambda \quad \text{for} \quad i = 1, \cdots, n.$$

Lemma 5. If T is *c.c.*, then the space TR can be topologized so that it becomes bicompact.

Proof. We can introduce a metric in TR by the following

$$\rho(T\eta_1, T\eta_2) = \sup_{\xi_F} |(\xi_F, T\eta_1) - (\xi^F, T\eta_2)| . \tag{23}$$

Then the condition of complete continuity of T implies that the space TR is compact by the metric (23). Accordingly the space TR is bicompact.

Lemma 6. If T is *c.c.*, then we have

$$\sup_\xi \inf_\eta (\xi, T\eta) = \lambda_0 . \tag{24}$$

Proof. If we show that, for any $\varepsilon > 0$, there exists $\xi \in P$ with

$$(\xi, T\eta) \geq \lambda_0 - \varepsilon \qquad \text{for all } \eta \in R \tag{25}$$

then Lemma 6 will be proved. Accordingly it is sufficient to prove (25) Let us well order the elements of P

$$\xi_1, \xi_2, \cdots, \xi_\nu, \cdots$$

We define subsets of TR as follows

$$(26)\quad G[\xi_\nu, \cdots, \xi_{\nu_n}] = \text{closure}\,[T\eta\,;\,(\xi_{\nu_i}, T\eta) < \lambda_0 - \varepsilon,\ i = 1, \cdots, n],$$

and

$$(27)\qquad G(\xi_\nu) = \Pi\, G[\xi_\nu, \xi_{\nu_1}, \cdots, \xi_{n_\nu}]$$

where the intersection is taken over all finite subsets with $\nu_1, \cdots, \nu_n$ less than ν. Then it is clear that

$$(28)\qquad G(\xi_1) \supset G(\xi_2) \supset \cdots \supset G(\xi_\nu) \supset \cdots \supset G(\xi_\mu) \supset \cdots$$

with $\nu < \mu$, and every $G(\xi_\nu)$ is a closed subset of TR. If (25) does not hold, then we shall prove that each $G(\xi_\nu)$ is not empty. For that let us assume that $G(\xi_\nu)$ is empty for a certain ν. Then, since by Lemma 5, the space TR is bicompact, there exists a finite product with

$$(29)\qquad 0 = \prod_{k=1}^{m} G[\xi_\nu, \xi_{\nu_{k1}}, \cdots, \xi_{\nu_{k n_k}}]$$
$$= G[\xi_\nu, \xi_{\nu_{11}}, \cdots, \xi_{\nu_{1 n_1}}, \cdots, \xi_{\nu_{m1}}, \cdots, \xi_{\nu_{m n_m}}]$$

Now, if (25) does not hold, then for every $\xi \in P$, there exists $\eta \in R$ with

$$(\xi, T\eta) < \lambda_0 - \varepsilon\,.$$

Therefore, by Lemma 4, (29) does not hold. This contradiction proves that each $G(\xi_\nu)$ is not empty.
Then, by the bicompactness of TR, there exists a $T\eta_0 \in TR$ which is common to the every member of (28), and hence

$$(\xi, T\eta_0) \leq \lambda_0 - \varepsilon \qquad \text{for every } \xi \in P.$$

Accordingly we have

$$(30)\qquad \sup_{\xi_F} (\xi_F\,;\, T\eta_0) < \lambda_0\,,$$

and this contradicts the definition of λ_0. Therefore (25) is proved.

Theorem 1. If $r(F, d)$ is a bounded function of F, d and for any fixed F, as a function of d, measurable B, and if T is *c.c.*, then the decision problem is strictly determined. In fact we have

$$(31)\qquad \lambda_0 = \min_{\eta\in R} \sup_{\xi\in P} r(\xi, \eta) = \sup_{\xi\in P} \min_{\eta\in R} r(\xi, \eta)\,.$$

Accordingly in this case there exist a minimax (randomized) solution.

Proof. By Lemma 3, we have

$$(32)\qquad \lambda_0 = \min_{\eta} \sup_{F} r(F, \eta) = \min_{\eta} \sup_{\xi} r(\xi, \eta)\,.$$

It is well known that

$$(33)\qquad \min_{\eta} \sup_{\xi} r(\xi, \eta) \geq \sup_{\xi} \inf_{\eta} r(\xi, \eta)\,.$$

On the other hand, by Lemma 6, we have

$$\sup_{\xi} \inf_{\eta} r(\xi, \eta) \geq \lambda_0. \tag{34}$$

Therefore from (32), (33) and (34), we have

$$\min_{\eta} \sup_{\xi} r(\xi, \eta) = \sup_{\xi} \inf_{\eta} r(\xi, \eta). \tag{35}$$

Since T is *c.c.*, it can be easyly shown that on the right hand side of (35) we can replace $\inf_{\eta}$ by $\min_{\eta}$, and Theorem 1 is proved.

Lemma 7. If the space D is compact with respect to the metric (1), then the operator T is *c.c.*.

Proof. Since this is essentially equivalent to Theorem 2.14 and 2.15 of [2], we shall omit the proof here.
From Theorem 1 and Lemma 7, it follows.

Theorem 2. If $r(F, d)$ is a bounded function of F, d and, for any fixed F, as a function of d measurable B, and if D is compact, then the decision problem is strictly determined. In fact we have (31).

In order to strengthen above theorems, we assume that, for any real number α and for any element $d \in D$, the multiplication αd is defined, and that for any two elements d_1 and d_2 of D the addition $d_1 + d_2$ is defined. Then we have the following

Theorem 3. If $r(F, d)$ is a bounded function of F, d and a covex, measurable B function of d for any fixed F, and if D is compact and convex, that is for any two real numbers α_1 and α_2 with $0 \leq \alpha_1 = 1 - \alpha_2 \leq 1$, and for any two elements d_1 and d_2 of D, $\alpha_1 d_1 + \alpha_2$ also belongs to D, then the decision problem is strictly determined. In fact we have

$$\min_{d \in D} \sup_{\xi \in P} r(\xi, d) = \sup_{\xi \in P} \min_{d \in D} r(\xi, d) \tag{36}$$

$$= \min_{\eta \in R} \sup_{\xi \in P} r(\xi, \eta) = \sup_{\xi \in O} \min_{\eta \in R} r(\xi, \eta). \tag{37}$$

Accordingly, in this case, there exists a non-randomized minimax solution. It is determined as a *d.f.* d_0 which minimizes $\sup_{F} r(F, d)$ with respect to d of D.

Proof. At first, under the conditions of Theorem 3, we shall prove that

$$\inf_{\eta} \sup_{F} r(F, \eta) = \inf_{d} \sup_{F} r(F, d). \tag{38}$$

Since $r(F, d)$ is bounded and D is compact, we can choose a sequence $\{\Delta_n\}$ of subdivisions of D, where

$$\Delta_n = \{D_{n_1}, \cdots, D_{n m_n}\}$$

D_{ni} and D_{nj}, $i \neq j$ are disjoint, and $D = \sum_{j=1}^{m_n} D_{nj}$, and an element d_{nj} of D_{nj} such that

$$(39) \qquad r(F, \eta) = \int r(F, d)\, d\eta = \lim_{n\to\infty} \sum_{j=1}^{m_n} r(F, d_{nj})\, \eta(D_{nj}).$$

If we write $\eta(D_{nj}) = \eta_{nj}$, then, since D is convex, we have

$$(40) \qquad d_n^* = \sum_j \eta_{nj} d_{nj} \in D.$$

And since $r(F, d)$ is a convex function of d, we have

$$(41) \qquad \sum_j r(F, d_{nj})\, \eta_{nj} \geq r(F, d_n^*).$$

Now, since D is compact with respect to the metric (1), we can choose a subsequence $\{d_{n_k}^*\}$ of $\{d_n^*\}$ and an element d^* of D with

$$(42) \qquad \lim_{k\to\infty} r(F, d_{n_k}^*) = r(F, d^*) \qquad \text{uniformly in } F.$$

Therefore from (39), (41) and (42) we have

$$(43) \qquad r(F, \eta) \geq r(F, d^*) \qquad \text{for every } F \in \Omega.$$

Then it follows

$$(44) \qquad \sup_F r(F, \eta) \geq \sup_F r(F, d^*) \geq \inf_d \sup_F r(F, d).$$

Since (44) holds for every $\eta \in R$, we have

$$(45) \qquad \inf_\eta \sup_F r(F, \eta) \geq \inf_d \sup_F r(F, d).$$

On the othere hand it is clear that in (45) $\leq$ holds. Therfore (38) is proved. From (38), it is clear that

$$(46) \qquad \inf_\eta \sup_\xi r(\xi, \eta) = \inf_d \sup_\xi r(\xi, d).$$

Now, it is clear that

$$(47) \qquad \sup_\xi \inf_\eta r(\xi, \eta) = \sup_\xi \inf_d r(\xi, d).$$

Since, by Theorem 2, the members of the left hand sides of (46) and (47) are equal, we have

$$(48) \qquad \inf_\eta \sup_\xi r(\xi, \eta) = \sup_\xi \inf_\eta r(\xi, \eta)$$

$$(49) \qquad = \inf_d \sup_\xi r(\xi, d) = \sup_\xi \inf_d r(\xi, d).$$

We can easily show that $\sup_\xi r(\xi, d)$ and $r(\xi, d)$ are continuous function of d. Accordingly, by the compactness of D, we can replace $\inf_d$ of both sides of (49) by $\min_d$, and the theorem is proved.

§ 3. Applications of the general Theory

We assume that the population distribution function is determined by the values of k parameters $\theta_1, \cdots, \theta_k$, and concerning these parameters we know that

$$(50) \qquad a_i \leq \theta_i \leq b_i \qquad (i = 1, 2, \cdots, k)$$

where a_i and b_i are given constants $(i = 1, \cdots, k)$.
Then we assume that it is requered to estimate one of the parameters, say θ_1, on the basis of a sample $x = (x_1, \cdots, x_n)$ of size n. (For the case, where more than one parameter are desired to estimate, the following procedure will also be applied.) For that we consider only these decision functions d which are given as the polynomial of $x_1, \cdots, x_n$ of the given type

$$(51) \qquad d(x) = \sum a_{m_1 \cdots m_n} x_1^{m_1} \cdots x_n^{m_n}$$

and in the range of these *d.f.*'s, we search for the minimax solution. If the given polynomial (51), definig a d.f. d, is the sum of N terms, then let its coefficients be $a_1, \cdots, a_N$, following the fixed order of the terms. Thenwe can represent a *d.f.* d as a point

$$d = (a_1, \cdots, a_N)$$

of the N dimensional Euclidian space E_N.
For any real number α and any *d.f.* d, we define αd as a *d.f.* defined by the point

$$\alpha d = (\alpha a_1, \cdots, \alpha a_N).$$

For any two *d.f.*'s d_1, d_2, such that

$$d_1 = (a_{11}, \cdots, a_{1N}), \; d_2 = (a_{21}, \cdots, a_{2N})$$

we define $d_1 + d_2$ as the d.f. defined by the point

$$d_1 + d_2 = (a_{11} + a_{21}, \cdots, a_{1N} + a_{2N}).$$

Let D^* be the subset of E_N consisting of all points which correspond to *d.f.*'s of D.

We assume that D^* is a convex, closed subset of E_N.
Let $W(\theta_1, \cdots, \theta_k; d(x))$ be the loss function, when $\theta_1, \cdots, \theta_k$ are true parameter values and we estimate θ_1 by $d(x)$ on the basis of a d.f. d and a sample x. Then we assume that $W(\theta_1, \cdots, \theta_k; d(x))$ is given by

$$(52) \qquad W(\theta_1, \cdots, \theta_k; d(x)) = (\theta_1 - d(x))^2 \cdot G(\theta_2, \cdots, \theta_k),$$

where $G(\theta_2, \cdots, \theta_k)$ is the given continious function of $\theta_2, \cdots, \theta_k$ in the range of (50). Then it can be easily shown that the conditions of Theorem 3 are satisfied. Accordingly, in this case, the decision problem is strictly determined and the minimax solution is detemrined as the non-randomized d.f. d_0 which minimizes $\max_{\theta_1 \cdots \theta_k} r(\theta_1, \cdots, \theta_k; d)$, where

(53) $$r(\theta_1, \cdots, \theta_k; d) = E[W(\theta_1, \cdots, \theta_k; d(x))].$$

We shall apply these results to the following problems.

Theorem 1. Minimax estimation of the mean.

We assume that the population distribution function is determined by the mean θ and the variance σ^2 which are independent algebraicllay, and it is known that

(54) $$-\alpha \leq \theta \leq \alpha \quad \text{and} \quad L^2 \leq \sigma^2 \leq K^2,$$

where α, L^2 and K^2 are given positive numbers.
Then, to estimate the mean θ on the basis of a sample $x = (x_1, \cdots, x_n)$ of size n, we assume that we consider only *d.f.*'s d of the type

(55) $$d(x) = \sum_{i=1}^{n} a_i x_i + b$$

and in the range of these d.f.'s we serach for the minimax solution. Then, if the loss function is given by

(56) $$W(\theta, \sigma^2; d(x)) = (\sum a_i x_i + b - \theta)^2,$$

then, the minimax solution is given with

(57) $$a_1 = \cdots = a_n = \frac{\alpha^2}{n\alpha^2 + K^2} \quad \text{and } b = 0$$

that is, the minimax solution d_0 is determined as follows

(58) $$d_0(x) = \frac{\alpha^2}{n\alpha^2 + K^2} \sum_{i=1}^{n} x_i .$$

If, in the above problem, the loss function is given by

(59) $$W(\theta, \sigma^2; d(x)) = \frac{1}{\sigma^2} (\sum a_i x_i + b - \theta)^2$$

then the minimax solution d_0 is determined as follows

(60) $$d_0(x) = \frac{\alpha^2}{n\alpha^2 + L^2} \sum_{i=}^{n} x .$$

Remark 1. When we say the minimax solution in Theorem 1, it means that, how the convex, closed subset of E_{n+1} consisting of points $(a_1, \cdots, a_n, b)$ corresponding to *d.f.*'s, may be large, the minimax solution in that range is given by (58) or (60). The same remark should be hold in all the following theorems.

Remark 2. If we have no information concerning the mean θ, then, in (53), α will be ∞. In this case we see that $d_0(x)$ of (58) and (60) both tend to $1/n \sum x_i$ as α tends to ∞.

Proof. At first we consider the case where the loss function is given by (56). Then we have

$$(61)\qquad r(\theta, \sigma^2; d) = E(\sum a_i x_i + b - \theta)^2$$
$$= \sigma^2 \sum_{i=1}^{n} a_i^2 + [(\sum a_i - 1)\theta + b]^2 .$$

Case I. When $\sum a_i - 1 \neq 0.$
If

$$(\text{I}, 1)\qquad \frac{-b}{\sum a_i - 1} \leq 0,$$

then we have

$$(62)\qquad \max_{\theta, \sigma^2} r(\theta, \sigma^2; d) = r(\alpha, K^2; d)$$
$$= [(\sum a_i - 1)\alpha + b]^2 + K^2 \sum a_i^2 \quad (= \varphi, \text{ say})$$

If

$$(\text{I}, 2)\qquad \frac{-b}{\sum a_i - 1} > 0$$

then we have

$$(63)\qquad \max_{\theta, \sigma^2} r(\theta, \sigma^2; d) = r(-\alpha, K^2; d)$$
$$= [-(\sum a_i - 1)\alpha + b]^2 + K^2 \sum a_i^2 \ (= \psi, \text{ say})$$

Case II. when $\sum a_i - 1 = 0.$
Then we have

$$(64)\qquad \max_{\theta, \sigma^2} r(\theta, \sigma^2; d) = b^2 + K^2 \sum a_i^2 \quad (= f, \text{ say}).$$

Now, we shall search for the values of $a_1, \cdots, a_n$ which minimizes sup r, b being fixed.
When the case (I, 1) holds, solving the equations

$$\frac{\partial \varphi}{\partial a_i} = 0 \qquad (i = 1, \cdots, n)$$

we have

$$a_1 = \cdots = a_n = \frac{\alpha^2 - \alpha b}{n\alpha^2 + K^2}.$$

Then we have

$$(65)\qquad \min_{a_1 \cdots a_n} \max_{\theta, \sigma^2} r(\theta, \sigma^2; d) = \frac{K^2(\alpha - b)^2}{n\alpha^2 + K^2} \qquad (= g(b), \text{ say}).$$

Now, the conditions of I and (I, 1) may be combined as follows

(66) $$-\frac{K^2}{n\alpha^2} < b \leq 0 .$$

Therefore, from the form of $g(b)$ and (66), we have

(67) $$\min_{d\in(\mathrm{I},1)} \max_{\theta,\sigma^2} r(\theta, \sigma^2; d) = \min_{b} g(b) = g(0) = \frac{\alpha^2 K^2}{n\alpha^2 + K^2},$$

where $\min_{d\in(\mathrm{I},1)}$ means the minimum value with respected to *d.f.*'s which satisfy the condition of (I, 1), and the minimum value is obtained when

(68) $$a_1 = \cdots = a_n = \frac{\alpha^2 K^2}{n\alpha^2 + K^2} \quad \text{and } b = 0.$$

For case (I, 2), we can show, in the same way as above, the followings

(69) $$\min_{d\in(\mathrm{I},2)} \max_{\theta,\sigma^2} r(\theta, \sigma^2; d) = \frac{\alpha^2 K^2}{n\alpha^2 + K^2}$$

where $\min_{v\in(\mathrm{I},2)}$ means the minimum value with respect to *d.f.*'s which statisfy the condition of (I, 2).

When the case (II) holds, we have from (64)

(70) $$\min_{a_1\cdots a_n, b} f = \min_{a_1\cdots a_n} K^2 \cdot \sum a_i{}^2 .$$

And since $\sum a_i = 1$ in this case, from (70), we have

(71) $$\min_{d\in(\mathrm{II})} \max_{\theta,\sigma_2} r(\theta, \sigma^2; d) = \frac{K^2}{n},$$

where $\min_{d\in(\mathrm{II})}$ means the minmum value with respect to *d.f.*'s which satisfy the condition of (II), and the above minimum value is obtained when

(72) $$a_1 = \cdots = a_n = \frac{1}{n} \quad \text{and } b = 0.$$

It is clear that

(73) $$\frac{K^2}{n} > \frac{\alpha^2 K^2}{n\alpha^2 + K^2}.$$

Therefore, from (67), (69), (71) and (73), if d_0 is the d.f. determined with (68), then we have

$$\max_{\theta,\sigma^2} r(\theta, \sigma^2; d_0) = \min_{d} \max_{\theta,\sigma^2} r(\theta, \sigma^2; d) = \frac{\alpha^2 K^2}{n\alpha^2 + K^2}.$$

Accordingly the d.f. d_0 is the minimax solution of this problem and the first half of Theorem 1 is proved.

For the case where the loss function is give by (59), the second half of Theorem 1 can be proved in the same way as above.

Remark 3. Here let us compare the risk fnnction of the minimax solution d_0 and the *d.f.* d^* which is given by

$$d^*(x) = \frac{1}{n} \sum_{i=1}^n x_i .$$

When the loss fuction is given by (56), it can be shown that

$$(74) \qquad r(\theta, \sigma^2 ; d_0) = \frac{1}{(n\alpha^2 + K^2)^2} [n\alpha^4 \sigma^2 + K^4 \theta^2]$$

and

$$(75) \qquad r(\theta, \sigma^2 ; d^*) = \sigma^2/n .$$

Therefore it can be shown that in the interior of the parabola

$$(76) \qquad \sigma^2 = \frac{nK^2}{2n\alpha^2 + K^2} \theta^2$$

on the plane of θ and σ^2, we have

$$r(\theta, \sigma^2 ; d^*) > r(\theta, \sigma^2 ; d_0)$$

and in the outside of the parabola (76), we have

$$r(\theta, \sigma^2 ; d^*) < r(\theta, \sigma^2 ; d_0) .$$

When the loss fucntion is given by (59), it can be shown that

$$(77) \qquad r(\theta, \sigma^2 ; d_0) = \frac{\theta^4}{\sigma^2} \cdot \frac{L^4}{(n\alpha^2 + L^2)^2} + \frac{n\alpha^4}{(n\alpha^2 + L^2)^2}$$

and

$$(78) \qquad r(\theta, \sigma^2 ; d^*) = \frac{1}{n} .$$

Therefore in the range of θ and σ^2 defined by (54), it can be shown that

$$r(\theta, \sigma^2 ; d^*) > r(\theta, \sigma^2 ; d_0) .$$

Theorem 2. Minimax prediction.

We assume that the population distribution function is determined by the mean θ and the variance σ^2 which are independent algebraically, and it is known that

$$(79) \qquad -\alpha \leq \theta \leq \alpha , \quad L^2 \leq \sigma^2 \leq K^2 ,$$

where α, L^2 and K^2 are given positive numbers.

We assume that when a sample $x = (x_1, \cdots, x_n)$ of size n is given, it is required to predict the mean of a sample $y = (y_1, \cdots, y_m)$ of size m which will be taken independently of the sample x, from the same population as was taken the sample x. Then we consider *d.f.*'s d of the type

$$(80) \qquad d(x) = \sum_{i=1}^n a_i x_i + b$$

and we shall search for the minimax solution in the range these *d.f.*'. If the risk function is given by

$$(81)\qquad r(\theta, \sigma^2; d) = E\left[(\sum_{i=1}^{n} a_i x_i + b) - \frac{1}{m}\sum_{j=1}^{m} y_j\right]^2$$

then the minmax solution d_0 is given by

$$(82)\qquad d_0(x) = \frac{\alpha^2}{n\alpha^2 + K^2} \sum_{i=1}^{n} x_i .$$

Proof. In this case, from (81), we have

$$(83)\qquad r(\theta, \sigma^2; d) = [(\sum a_i - 1)\theta + b]^2 + \sigma^2 [\sum a_i^2 + 1/m] .$$

Therefore compairing (83) with (61), it can be seen that the minimax solution of this problem is given by (82) as in Theorem 1.

Theorem 3. Minimax estimation of the variance.

For the normal population, we assume that the mean value is known to be zero, and concerning the variance σ^2, we know that

$$(84)\qquad L^2 \leq \sigma^2 \leq K^2 ,$$

where L^2 and K^2 are given positive numbers.
Then to estimate the variance σ^2 on the basis of a sample $x = (x_1, \cdots, x_n)$ of size n, we consider only *d.f.*'s of the type

$$(85)\qquad d(x) = \sum_{i=1}^{n} a_i x_i^2 + \sum_{i,j} b_{ij} x_i x_j + \sum_{i=1}^{n} c_i x + 1, \quad (b_{ij} = b_{ji},\ i \neq j)$$

and in the range of these *d.f*'s we shall search for the minimax solution. If the loss function is given by

$$(86)\qquad W(\sigma^2; d(x)) = (d(x) - \sigma^2)^2 .$$

Then the minimax solution d_0 is determined as follows

$$(87)\qquad d_0(x) = \frac{1}{n+2} \sum_{i=1}^{n} x_i^2 .$$

Proof. In this case, it can be shown that

$$(88)\qquad \begin{aligned} r(\sigma^2; d) &= E[W(\sigma^2, d(x))] \\ &= \sigma^4 [3\sum_i a_i^2 + \sum_{i,j} a_i a_j + \sum_{i,j} b_{ij}^2 - 2\sum_i a_i + 1] \\ &\quad + \sigma^2 [\sum_i c_i^2 - 2l + 2l \sum_i a_i] + l^2 . \end{aligned}$$

We put this as follows

$$(89)\qquad r(\sigma^2; d) = A\sigma^4 + B\sigma^2 + C .$$

Then, it is clear that $A > 0$.

Case I. when

$$-\frac{B}{2A} \leq \frac{L^2 + K^2}{2}, \quad \text{that is,} \quad -B \leq (L^2 + K^2) A.$$

Then, from (89), it is clear that

$$\sup_{\sigma^2} r(\sigma^2 ; d) = r(K^2 ; d) . \tag{90}$$

Case II. when

$$-\frac{B}{2A} > \frac{L^2 + K^2}{2}, \quad \text{that is,} \quad -B > (L^2 + K^2) A .$$

Then, from (89), it is clear that

$$\sup_{\sigma^2} r(\sigma^2 ; d) = r(L^2 ; d) . \tag{91}$$

Now, for any fixed value of σ^2, we search for values of a's, b's, c's and l, which minimizes $r(\sigma^2 ; d)$ with respect to d. Then, as the solutions of equations

$$\frac{\partial r}{\partial a_i} = 0, \quad \frac{\partial r}{\partial b_{ij}} = 0, \quad \frac{\partial r}{\partial c_i} = 0, \quad \frac{\partial r}{\partial l} = 0 \tag{92}$$
$$i, j = 1, \cdots, n.$$

We have

$$a_1 = a_2 = \cdots = a_n = \frac{1}{n+2} \tag{93}$$
$$b_{ij} = 0, \quad c_i = 0, \quad l = 0,$$
$$i, j = 1, \cdots, n.$$

Here we remark that this solution is independent of the fixed value of σ^2. It is clear that values of a's, b's, c's and l given by (93) satisfy the condition of the case I. Therefore we have

$$\min_{d \in [I]} \max_{\sigma^2} r(\sigma^2 ; d) = \min_{d \in [I]} r(K^2 ; d) = r(K^2 ; d_0) = \frac{2}{n+2} K^4, \tag{94}$$

where $\min\limits_{d \in [I]}$ shows the minimum value with respect to $d.f.$'s which satisfy the condition of the case I, and $d_0(x)$ is the $d.f.$ determined by (93), that is

$$d_0(x) = \frac{1}{n+2} \sum_{i=1}^{n} x_i^2 . \tag{95}$$

Next, let d^* be any fixed $d.f.$ which satisfies the condition of the case II, then we have from (91)

$$\sup_{\sigma^2} r(\sigma^2 ; d^*) = r(L^2 ; d^*) . \tag{96}$$

Therefore we have

$$(97) \qquad r(K^2;\, d^*) \leq r(L^2;\, d^*),$$

and from the above remark,

$$(98) \qquad r(K^2;\, d_0) \leq r(K^2; d^*).$$

Accordingly, from (96) – (98), we have

$$(99) \qquad r(K^2;\, d_0) \leq \sup_{\sigma^2} r(\sigma^2;\, d^*)$$

for any *d.f.* d^* which satisfies the condition of the case II. From (94) and (99), we have

$$(100) \qquad \min_{d\in[\mathrm{I}]} \max_{\sigma^2} r(\sigma^2;\, d) \leq \inf_{d\in[\mathrm{II}]} \max_{\sigma^2} r(\sigma^2;\, d).$$

Accordingly from (100), we have

$$\min_{d} \max_{\sigma^2} r(\sigma^2;\, d) = r(K^2;\, d_0) = \frac{2}{n+2} K^4.$$

Therefore, d_0 is the minimax solution of this problem and Theorem 3 is proved.

Remark 4. Let d^* be the d.f. given by

$$d(x) = \frac{1}{n} \sum_{i=1}^{n} x_i^2,$$

then it can be shown easyly that

$$r(\sigma^2;\, d_0) = \frac{2}{n+2}\sigma^4 < r(\sigma^2;\, d^*) = \frac{2}{n}\sigma^4.$$

Theorem 4. Minimax estimation in the binomial population.
Let X be the random variable which represents the number of occurences of the event A in n independent trials, each having the probability θ of occurence of the event A. Then to estimate θ on the basis of the observed value x of X, we consider only *d.f.*'s of the type

$$(101) \qquad d(x) = ax + b,$$

and in the range of these *d.f.*'s we search for the minimax solution.
If the loss function is given by

$$(102) \qquad W(\theta, d(x)) = (\theta - d(x))^2$$

then the minimax solution d_0 is given by

$$(103) \qquad d_0(x) = \frac{1}{\sqrt{n}+1}\left(\frac{x}{\sqrt{n}} + \frac{1}{2}\right).$$

Remark 5. This result was alredy obtained in [3], using an apropriate a priori distribution which is taken intuitively. But here we shall solve the problem directly following the above method.

Proof. Proving this theorem we shall omit complicated calculations and state only the key points. We have in this case

$$(104)\quad r(\theta; d) = E(ax + b - \theta)^2$$
$$= [n(n-1)a^2 - 2na + 1]\theta^2 + [na^2 + 2b(na-1)]\theta + b^2$$

The roots of the equation

$$(105)\qquad n(n-1)a^2 - 2na + 1 = 0$$

are given by

$$(106)\qquad a_1 = \frac{1}{\sqrt{n}(\sqrt{n}+1)},\quad a_2 = \frac{1}{\sqrt{n}(\sqrt{n}-1)}.$$

It can be shown under the condition

$$(107)\qquad a < a_1$$

we have

$$(108)\qquad \inf_d \sup_\theta r(\theta; d) = \frac{1}{4(\sqrt{n}+1)^2},$$

where $\inf_d$ is the infimum with respect to *d.f.*'s which satisfy the condition (107).
Under the condition

$$(109)\qquad a > a_2$$

we have

$$(110)\qquad \inf_d \sup_\theta r(\theta; d) = \frac{1}{4(\sqrt{n}-1)^2},$$

where $\inf_d$ is the infimum with respect to *d.f.*'s which satisfy the condition (109).
Under the condition

$$(111)\qquad a_1 < a < a_2,$$

we have

$$(112)\qquad \inf_d \sup_\theta r(\theta; d) = \frac{1}{4(\sqrt{n}+1)^2},$$

where $\inf_d$ is the infimum with respect to *d.f.*'s which satisfy the condition (111).
Under the condition

$$(113)\qquad a = a_1$$

we have

$$(114)\qquad \min_d \sup_\theta r(\theta; d) = \frac{1}{4(\sqrt{u}+1)^2},$$

where $\min_d$ is the minimum with respect to *d.f.*'s which satisfy the condition (113), and this minimum value is attained with

$$(115) \qquad a = a_1 \quad \text{and} \quad b = \frac{1}{2(\sqrt{n}+1)}.$$

Under the condition

$$(116) \qquad a = a_2$$

we have

$$(117) \qquad \min_d \sup_\theta r(\theta\,;d) = \frac{1}{4(\sqrt{n}-1)^2},$$

where $\min_d$ is the minimum with respect to *d.f.*'s which satisfy the condition (116).

Therefore, from (108), (110), (112), (114) and (117), we conclude that

$$(118) \qquad \min_d \sup_\theta r(\theta\,;d) = \sup_\theta r(\theta\,;d_0) = \frac{1}{4(\sqrt{n}+1)^2},$$

where $\min_d$ is the minimum with respect to *d.f.*'s of the type (101), taking account of Remark 1, and d_0 is the *d.f.* given by

$$a = a_1 = \frac{1}{\sqrt{n}(\sqrt{n}+1)} \quad \text{and,} \quad b = \frac{1}{2(\sqrt{n}+1)}.$$

Therefor the d.f. d_0 such that

$$d_0(x) = \frac{1}{\sqrt{n}+1}\left(\frac{x}{\sqrt{n}} + \frac{1}{2}\right)$$

is the minimax solution of this problem and Theorem 4 is proved.

References

[1] S. Karlin, "Operator treatment of Minimax Principle". Contributions to the Theory of Games, Princeton University Press, 1950.

[2] A. Wald, "Statistical Decision Functions", 1950.

[3] J. L. Hodges, E. L. Lehmann, "Some Problems in minimax point estimation" Ann. Math. Stat. vol. 21 (1950).