Randomized minimax estimators under simple random sampling from a finite population

Randomized minimax estimators under simple random sampling from a finite population

Abstract. The unknown number of items within a finite population which have a certain property is to be estimated after drawing a sample without replacement. In case that the estimates are allowed to be arbitrary reals and that squared error loss is assumed, an explicit formula for the minimax estimator has already been determined by Hodges and Lehmann in 1950. But if the analysis is restricted to integer valued estimators such a neat solution of the minimax problem under squared error loss is not at hand. To ensure that the corresponding statistical game is strictly determined randomized integer valued estimators are considered in this paper, and sufficient conditions are derived for a randomized estimator to be minimax as well as for a prior to be least favourable. Numerical results are presented at the end of the paper.

1. Notation and introduction

Consider a finite population of N items, θ of which have a certain property. The unknown frequency θ, which is an element of the parameter set $\Theta = \{0, 1, \dots, N\}$, is to be estimated after a simple random sample (i.e. without replacement) of size n has been drawn. The number of items in the sample which have the specified property is a sufficient statistic having a hypergeometric distribution. Therefore $\mathbf{X} = \{0, 1, \dots, n\}$ is an appropriate sample space. In order to avoid trivial cases it is assumed that $n \leq N - 1$. An elementary and detailed description of the following decision theoretic framework is given in [4]. The notation used below is basicly in accordance with that in [1]. Let Δ be the set of all randomized estimators, i.e. the set of all $n+1$-tupels $\delta = (\delta_0, \dots, \delta_n)$ of probability measures

$$\delta_x = \sum_{a=0}^{N} \alpha_{xa} \cdot \varepsilon_a, \qquad x \in \mathbf{X}, \tag{1}$$

on the action space $A = \{0, 1, \dots, N\}$ where $\alpha_{x0}, \dots, \alpha_{xN} \geq 0$ and $\alpha_{x0} + \dots + \alpha_{xN} = 1$ for $x \in \mathbf{X}$ and ε_a denotes the one-point measure which puts its mass on a. Let Π be the set of all priors, i.e. the set of all probability measures

$$\pi = \sum_{\theta=0}^{N} p_\theta \cdot \varepsilon_\theta \tag{2}$$

on the parameter space Θ where $p_0, \dots, p_N \geq 0$ and $p_0 + \dots + p_N = 1$. The Bayes risk of a randomized estimator δ with respect to a prior π according to (1) and (2), respectively, is defined by

$$r(\pi, \delta) = \sum_{\theta=0}^{N} R(\theta, \delta)\, p_\theta$$

where $R(\cdot,\delta)$ denotes the risk function of δ given by

$$R(\theta,\delta)=\frac{1}{\binom{N}{n}}\sum_{x=0}^{n}\sum_{a=0}^{N}\alpha_{xa}(\theta-a)^2\binom{\theta}{x}\binom{N-\theta}{n-x},\qquad \theta\in\Theta, \tag{3}$$

under squared error loss. A randomized estimator δ_π with

$$r(\pi,\delta_\pi)=\inf_{\delta\in\Delta} r(\pi,\delta)$$

is called Bayes with respect to the prior π. The minimax risk r^* is defined by

$$r^*=\inf_{\delta\in\Delta}\ \sup_{\pi\in\Pi} r(\pi,\delta), \tag{4}$$

and a randomized estimator δ^* with

$$\sup_{\pi\in\Pi} r(\pi,\delta^*)=r^*$$

is called minimax. A prior π^* with

$$\inf_{\delta\in\Delta} r(\pi^*,\delta)=\sup_{\pi\in\Pi}\ \inf_{\delta\in\Delta} r(\pi,\delta)$$

is called least favourable, and the statistical game (Π,Δ,r) is said to be strictly determined if

$$\sup_{\pi\in\Pi}\ \inf_{\delta\in\Delta} r(\pi,\delta)=r^*.$$

The following result is well known (see e.g. [6], Theorem 3.20): The statistical game (Π,Δ,r) is strictly determined, π^* is a least favourable prior, and δ^* is a minimax estimator if and only if (π^*,δ^*) is a saddle-point in (Π,Δ,r), i.e. if and only if

$$\inf_{\delta\in\Delta} r(\pi^*,\delta)=r(\pi^*,\delta^*)=\sup_{\pi\in\Pi} r(\pi,\delta^*).$$

Hodges and Lehmann consider the estimation problem as described above. The only difference is that they assume the action space A to be the set of reals. They prove (cf. [2], section 5, and [3], example 4.2.6) that the non-randomized estimator $\delta_{\mathbb{R}}$ defined by

$$\delta_{\mathbb{R}}(x)=N\cdot\left(x+\frac{1}{2}\sqrt{\frac{n(N-n)}{N-1}}\right)\cdot\left(n+\sqrt{\frac{n(N-n)}{N-1}}\right)^{-1},\qquad x\in\mathbf{X},$$

is minimax and that the minimax risk is given by

$$r_{\mathbb{R}}=\frac{N^2}{4}\cdot\frac{n(N-n)}{N-1}\cdot\left(n+\sqrt{\frac{n(N-n)}{N-1}}\right)^{-2}. \tag{5}$$

This estimator has the obvious disadvantage that the estimates $\delta_{\mathbb{R}}(x)$ are not necessarily integers whereas the unknown frequency θ, which is to be estimated, is one of the numbers $0, 1, \dots, N$. Therefore, it may be assumed that in applications the non-randomized estimator $\delta_{\mathbb{Z}}$ defined by

$$\delta_{\mathbb{Z}}(x) = [\delta_{\mathbb{R}}(x) + \tfrac{1}{2}], \qquad x \in \mathbf{X},$$

is used instead of the non-randomized estimator $\delta_{\mathbb{R}}$ where $[y]$ denotes the greatest integer less than or equal to y. However, Table 1 at the end of this paper shows that in most cases the maximum risk

$$r_{\mathbb{Z}} = \max_{\theta \in \Theta} R(\theta, \delta_{\mathbb{Z}}) \tag{6}$$

of this estimator is greater than the minimax risk r^* with respect to the adequate action space $A = \{0, 1, \dots, N\}$, i.e. the estimator $\delta_{\mathbb{Z}}$ is not minimax.

2. Computation of minimax estimators

The statistical game (Π, Δ, r) is the mixed extension of an appropriate finite game. Therefore it follows from a well-known result of John von Neumann [5] that there exists a saddle-point in (Π, Δ, r) which can be computed by linear programming techniques (cf. [6], Theorem 3.25 and ch. III.5). However, this approach causes great computational effort if the population is of reasonable size N (cf. [6], p. 206). In order to avoid this difficulty another method is described for solving the statistical game (Π, Δ, r). In the following theorem sufficient conditions are established for a prior π and a randomized estimator δ to form a saddle-point in (Π, Δ, r). It is rather simple to check whether these conditions are satisfied since this can be done by computing the solutions of two systems of $n+2$ linear equations. These systems are defined by

$$B \cdot \alpha = c, \qquad \alpha \in \mathbf{R}^{n+2}, \tag{7}$$

and

$$D \cdot p = e, \qquad p \in \mathbf{R}^{N+1}, \tag{8}$$

with matrices $B = (b_{ij})_{0 \le i, j \le n+1}$ and $D = (d_{ij})_{0 \le i \le n+1, 0 \le j \le N}$ as well as vectors $(c_i)_{0 \le i \le n+1}$ and $e = (e_i)_{0 \le i \le n+1}$ given by

$$b_{ij} = \begin{cases} (2i - 2a_j - 1)\dbinom{i}{j}\dbinom{N-i}{n-j} & \text{for } j \in \{0, 1, \dots, n\} \\ -\dbinom{N}{n} & \text{for } j = n+1 \end{cases}$$

for $i \in \{0, 1, \ldots, n+1\}$,

$$c_i = -\sum_{j=0}^{n} (i - a_j - 1)^2 \binom{i}{j} \binom{N-i}{n-j}$$

for $i \in \{0, 1, \ldots, n+1\}$,

$$d_{ij} = \begin{cases} (2j - 2a_i - \eta_i) \binom{j}{i} \binom{N-j}{n-i} & \text{for } i \in \{0, 1, \ldots, n\} \\ 1 & \text{for } i = n+1 \end{cases}$$

for $j \in \{0, 1, \ldots, N\}$, and

$$e_i = \begin{cases} 0 & \text{for } i \in \{0, 1, \ldots, n\} \\ 1 & \text{for } i = n+1 \end{cases}$$

where $a_0, \ldots, a_n$ and $\eta_0, \ldots, \eta_n$ are arbitrary elements of the sets $\{0, 1, \ldots, N-1\}$ and $\{0, 1, 2\}$, respectively.

Theorem. *Let $a_0, \ldots, a_n$ be elements of the set $\{0, 1, \ldots, N-1\}$ with the following two properties.*

(i) There exists a solution $\alpha = (\alpha_0, \ldots, \alpha_n, \varrho) \in [0, 1]^{n+1} \times \mathbf{R}$ of system (7).
(ii) There exists a solution $p = (p_0, \ldots, p_N) \in [0, 1]^{N+1}$ of system (8) with

$$\eta_i = \begin{cases} 0 & \text{for } \alpha_i = 1 \\ 1 & \text{for } \alpha_i \in (0, 1) \\ 2 & \text{for } \alpha_i = 0 \end{cases}$$

for $i \in \{0, 1, \ldots, n\}$.

Define a randomized estimator $\delta = (\delta_0, \ldots, \delta_n)$ by

$$\delta_x = \alpha_x \cdot \varepsilon_{a_x} + (1 - \alpha_x) \cdot \varepsilon_{a_x + 1}, \qquad x \in \mathbf{X}, \tag{9}$$

and a prior π by

$$\pi = \sum_{\theta=0}^{N} p_\theta \cdot \varepsilon_\theta \, .$$

Then (π, δ) is a saddle-point in the statistical game (Π, Δ, r), i.e.

(a) (Π, Δ, r) is strictly determined,
(b) the prior π is least favourable, and
(c) the randomized estimator δ is minimax.

The minimax risk is given by $r^* = \varrho$. The minimax estimator is uniquely determined in case that the matrix B is non-singular, $p_x + \ldots + p_{x+N-n} > 0$ for $x \in \mathbf{X}$, and $\# \{\theta \in \Theta \mid p_\theta > 0\} \geq n + 2$.

Proof. Subsequently it is shown that the randomized estimator δ is an equalizer rule and that it is Bayes with respect to the prior π. According to (3) the risk function of the randomized estimator δ is given by

$$R(\theta,\delta)=\frac{1}{\binom{N}{n}}\sum_{x=0}^{n}[\alpha_x(\theta-a_x)^2+(1-\alpha_x)(\theta-a_x-1)^2]\binom{\theta}{x}\binom{N-\theta}{n-x}$$

$$=\frac{1}{\binom{N}{n}}\sum_{x=0}^{n}[\alpha_x(2\theta-2a_x-1)+(\theta-a_x-1)^2]\binom{\theta}{x}\binom{N-\theta}{n-x}$$

for $\theta\in\Theta$. Hence it follows from the hypothesis (i) that $R(\theta,\delta)=\varrho$ for $\theta\in\{0,1,\dots,n+1\}$. This risk function has the form of a polynomial in θ of degree $n+1$ at most, and therefore it has to be constant. In particular it follows that $R(\theta,\delta)=\varrho$ for $\theta\in\Theta$, i.e. the randomized estimator δ is an equalizer rule.

Let $\tilde{\delta}$ be a randomized estimator, i.e.

$$\tilde{\delta}_x=\sum_{a=0}^{N}\alpha_{xa}\cdot\varepsilon_a,\qquad x\in\mathbf{X},$$

where $\alpha_{x0},\dots,\alpha_{xN}\geq 0$ and $\alpha_{x0}+\dots+\alpha_{xN}=1$ for $x\in\mathbf{X}$ according to (1). The Bayes risk of the randomized estimator $\tilde{\delta}$ with respect to the prior π can be written in the form

$$r(\pi,\tilde{\delta})=\frac{1}{\binom{N}{n}}\sum_{x=0}^{n}\sum_{a=0}^{N}\alpha_{xa}h_x(a)$$

where the function h_x is defined by

$$h_x(a)=a^2\cdot\sum_{\theta=0}^{N}\binom{\theta}{x}\binom{N-\theta}{n-x}p_\theta-2a\cdot\sum_{\theta=0}^{N}\theta\binom{\theta}{x}\binom{N-\theta}{n-x}p_\theta$$
$$+\sum_{\theta=0}^{N}\theta^2\binom{\theta}{x}\binom{N-\theta}{n-x}p_\theta,\qquad a\in A,$$

for $x\in\mathbf{X}$. Therefore the minimum Bayes risk of the prior π is given by

$$\inf_{\tilde{\delta}\in\Delta}r(\pi,\tilde{\delta})=\frac{1}{\binom{N}{n}}\sum_{x=0}^{n}\inf_{\substack{\alpha_{x0},\dots,\alpha_{xN}\geq 0\\ \alpha_{x0}+\dots+\alpha_{xN}=1}}\sum_{a=0}^{N}\alpha_{xa}h_x(a),$$

and a randomized estimator $\tilde{\delta}$ is Bayes with respect to the prior π if and only if the corresponding weights satisfy $\alpha_{xa}=0$ for $a\in A_x$ and $x\in\mathbf{X}$ where

$$A_x=\{\bar{a}\in A\,|\,h_x(\bar{a})>\min_{a\in A}h_x(a)\},\qquad x\in\mathbf{X},$$

denotes the set of all points which do not minimize the function h_x on A. Now, let $x \in \mathbf{X}$ be fixed. If

$$\sum_{\theta=0}^{N} \binom{\theta}{x} \binom{N-\theta}{n-x} p_\theta = 0,$$

i.e. if $p_x = \ldots = p_{x+N-n} = 0$, then $h_x(a) = 0$ for $a \in A$ and hence $A_x = \emptyset$. *If*

$$\sum_{\theta=0}^{N} \binom{\theta}{x} \binom{N-\theta}{n-x} p_\theta > 0,$$

i.e. if $p_x + \ldots + p_{x+N-n} > 0$, then the function h_x is a parabola which is minimized on $\mathbb{R}$ at

$$a_x^* = \frac{\sum_{\theta=0}^{N} \theta \binom{\theta}{x} \binom{N-\theta}{n-x} p_\theta}{\sum_{\theta=0}^{N} \binom{\theta}{x} \binom{N-\theta}{n-x} p_\theta} \in [0, N],$$

and hence

$$A_x = \begin{cases} A \backslash \{[a_x^* + \frac{1}{2}]\} & \text{for } a_x^* \notin \{\frac{1}{2}, \frac{3}{2}, \ldots, N - \frac{1}{2}\} \\ A \backslash \{a_x^* - \frac{1}{2},\ a_x^* + \frac{1}{2}\} & \text{for } a_x^* \in \{\frac{1}{2}, \frac{3}{2}, \ldots, N - \frac{1}{2}\} \end{cases}.$$

Therefore it follows from the hypothesis (ii) that the randomized estimator δ is Bayes with respect to the prior π which shows that (π, δ) is a saddle-point in the statistical game (Π, Δ, r) and that the minimax risk is given by $r^* = \varrho$.
Now assume additionally that the matrix B is non-singular, $p_x + \ldots + p_{x+N-n} > 0$ for $x \in \mathbf{X}$, and $\# \{\theta \in \Theta \,|\, p_\theta > 0\} \geq n + 2$. Let δ' be another minimax estimator. Since the statistical game (Π, Δ, r) is strictly determined it follows that (π, δ') is a saddle-point, too. Therefore the hypothesis $p_x + \ldots + p_{x+N-n} > 0$ for $x \in \mathbf{X}$ implies that δ' can be written in the form

$$\delta'_x = \alpha'_x \cdot \varepsilon_{a_x} + (1 - \alpha'_x) \cdot \varepsilon_{a_x+1}, \qquad x \in \mathbf{X},$$

for suitable weights $\alpha'_0, \ldots, \alpha'_n \in [0, 1]$. The risk function of the randomized estimator δ' satisfies $R(\theta, \delta') = \varrho$ for $\theta \in \Theta$ with $p_\theta > 0$. Since $R(\theta, \delta')$ has the form of a polynomial in θ of degree $n + 1$ at most it follows from the hypothesis $\# \{\theta \in \Theta \,|\, p_\theta > 0\} \geq n + 2$ that the randomized estimator δ' is an equalizer rule. Hence a short calculation yields $B \cdot \alpha' = c$ where $\alpha' = (\alpha'_0, \ldots, \alpha'_n, \varrho) \in [0, 1]^{n+1} \times \mathbb{R}$. Since the matrix B is assumed to be non-singular it follows that $\alpha = \alpha'$, and therefore $\delta = \delta'$. □

3. Numerical results

The theorem has been applied to determine a saddle-point in the statistical game (Π, Δ, r) for a population of $N = 10$ items and samples of size $n \in \{1, 2, \ldots, 9\}$. In each case the minimax estimator is uniquely determined whereas different least favourable priors exist for $n \leq 8$. Table 1 contains the minimax risks $r_{\mathbb{R}}$ and r^* according to (5) and (4), respectively, as well as the maximum risk $r_{\mathbb{Z}}$ according to (6). In Table 2 the parameters $a_0, \ldots, a_n$ and $\alpha_0, \ldots, \alpha_n$ of the uniquely determined minimax estimator according to (9) are given.

Table 1. Risks for $N = 10$.

n	$r_{\mathbb{R}}$	r^*	$r_{\mathbb{Z}}$
1	6.25000	6.50000	9.00000
2	4.00000	4.00000	4.00000
3	2.84574	3.00000	4.00000
4	2.10102	2.24528	2.66667
5	1.56250	1.75000	2.00000
6	1.14424	1.30709	1.60000
7	0.80218	1.00000	1.50000
8	0.51020	0.63353	1.00000
9	0.25000	0.50000	1.00000

Table 2. The minimax estimators for $N = 10$.

n	1	2	3	4	5	6	7	8	9
a_0	2	2	1	1	1	1	1	0	0
α_0	0.500	1.000	0.333	0.585	0.750	0.898	1.000	0.366	0.500
a_1	7	5	3	3	2	2	2	1	1
α_1	0.500	1.000	0.111	0.802	0.250	0.630	0.857	0.300	0.500
a_2		8	6	5	4	3	3	2	2
α_2		1.000	0.889	1.000	0.750	0.343	0.714	0.228	0.500
a_3			8	6	5	5	4	3	3
α_3			0.667	0.198	0.250	1.000	0.571	0.142	0.500
a_4				8	7	6	5	5	4
α_4				0.415	0.750	0.657	0.429	1.000	0.500
a_5					8	7	6	6	5
α_5					0.250	0.370	0.286	0.858	0.500
a_6						8	7	7	6
α_6						0.102	0.143	0.772	0.500
a_7							9	8	7
α_7							1.000	0.700	0.500
a_8								9	8
α_8								0.634	0.500
a_9									9
α_9									0.500

Acknowledgement. The authors would like to thank the Deutsche Forschungsgemeinschaft for financial support.

Jürgen Eichenauer, Jürgen Lehn
Fachbereich Mathematik der Technischen Hochschule Darmstadt

REFERENCES

1 Berger J. O.: Statistical decision theory and Bayesian analysis, 2nd ed., Springer, Berlin – Heidelberg – New York 1985.
2 Hodges J. L. and Lehmann E. L.: Some problems in minimax point estimation. Ann. Math. Statist. *21,* 182–197 (1950).
3 Lehmann E. L.: Theory of point estimation. Wiley, New York 1983.
4 Loeffel H.: Statistische Inferenz und strategisches Spiel. El. Math. *40,* 109–120 (1985).
5 Neumann J. v.: Zur Theorie der Gesellschaftsspiele. Math. Annalen *100,* 295–320 (1928).
6 Rauhut B., Schmitz N. und Zachow E.-W.: Spieltheorie. Teubner, Stuttgart 1979.

 0013-6018/88/060170-08$1.50+0.20/0

Some integral inequalities

The aim of this note is to prove some integral inequalities and to find interesting applications for the logarithmic and exponential functions. These relations have some known corollaries ([3], [4], [5], [8]).

Theorem 1. *Let $f:[a,b] \to \mathbb{R}$ $(a < b)$ be a differentiable function with increasing (strictly increasing) derivative on $[a,b]$. Then one has the following inequalities:*

$$\int_a^b f(t)\,dt \underset{(>)}{\geq} (b-a)\, f\left(\frac{a+b}{2}\right) \tag{1}$$

$$2 \cdot \int_a^b f(t)\,dt \underset{(<)}{\leq} (b-a)\, f(\sqrt{ab}) + (\sqrt{b} - \sqrt{a})(\sqrt{b}\, f(b) + \sqrt{a}\, f(a))$$

$$\textit{(Here } 0 \leq a < b\textit{).} \tag{2}$$

Proof. The Lagrange mean-value theorem implies: $f(y) - f(x) \underset{(>)}{\geq} (y-x)\, f'(x)$ for all $x, y \in [a,b]$. Take $x = (a+b)/2$ and integrate the obtained inequality:

$$\int_a^b f(y)\,dy - (b-a)\, f\left(\frac{a+b}{2}\right) \underset{(>)}{\geq} f'\left(\frac{a+b}{2}\right) \cdot \int_a^b \left(y - \frac{a+b}{2}\right) dy = 0,$$

i.e. relation (1).

In order to prove (2) consider as above the inequality $f(y) - f(x) \underset{(<)}{\leq} (y-x)\, f'(y)$ with $x = \sqrt{ab}$. Integrating by parts on $[a,b]$ we get

$$\int_a^b f(y)\,dy - (b-a)\, f(\sqrt{ab}) \underset{(<)}{\leq} (y - \sqrt{ab})\, f(y)\Big|_a^b - \int_a^b f(y)\,dy$$

which easily implies (2).