האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

INVARIANT AND MINIMAX ESTIMATION
OF QUANTILES IN FINITE POPULATIONS

By

YAAKOV MALINOVSKY and YOSEF RINOTT

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

BEST INVARIANT AND MINIMAX ESTIMATION OF QUANTILES IN FINITE POPULATIONS

By Yaakov Malinovsky[‡,*,†], Yosef Rinott[§,*]

National Institute of Child Health and Human Development[‡], The Hebrew University of Jerusalem, Department of Statistics and Center for the Study of Rationality.[§]

We study estimation of finite population quantiles, with emphasis on estimators that are invariant under monotone transformations of the data, and suitable invariant loss functions. We discuss non-randomized and randomized estimators, best invariant and minimax estimators and sampling strategies relative to different classes.

The combination of natural invariance of the kind discussed here, and finite population sampling appears to be novel, and leads to interesting statistical and combinatorial aspects.

1. Introduction. In this paper we study invariant estimation of quantiles of a finite population. While much of statistics, such as official statistics, concerns finite population sampling, with emphasis on estimation of totals and quantiles, most of the work in the past three decades or so on optimality properties of estimators, including the study of invariance, has concentrated on i.i.d sampling, that is, sampling from infinite populations.

In finite population sampling, the statistician chooses a *strategy* which consists of a *sampling design*, and an *estimator*, and the data consist of the labels of the sampled units, and their corresponding measured values; this clearly differs from infinite population sampling, where there is no sampling design to consider, and no labels.

When estimating quantiles of a finite population, it is natural to deal with estimators that are invariant under monotone transformations of the measured values, since under such transformations the population unit which represents the estimated quantile remains unchanged. It is also natural to consider the possibility of invariance under permutations of the labels.

1

Thus, in this paper we deal with best-invariant and minimax strategies, that is, sampling designs and estimators, for estimation of quantiles in connection with two groups. The infinite (and non-compact) group of monotone transformations, and the finite group of permutations. Another special aspect of the present work is that we consider a loss function which essentially measures the deviation of the estimate from the estimated quantile in terms of the number of population units that separate them; see (2.2). This invariant loss function has a combinatorial flavor, and so do some of our proofs, including that of Theorem 4.2 which is given in Malinovsky and Rinott (2009) and a simple use of the celebrated Ramsey Theorem in Theorem 5.3.

Some relevant references: invariance under monotone transformation when estimating a whole distribution function with various loss functions appears, for example, in Agarwal (1955), Ferguson (1967), Brown (1988), Yu and Chow (1991), Yu and Phadia (1992), Stępień–Baran (2009), Cohen and Kuo (1985), and Lehmann and Casella (1998), where the only last two reference consider finite population models. Invariant quantile estimation in infinite populations appears in Ferguson (1967), Brown (1988) (median) and Zieliński (1999).

Invariance in finite populations appears already in Blackwell and Girshick (1954), where only finite groups (permutations) are considered, and in many later references, such as Cassel et al. (1977), where invariance under linear transformations also appears.

General results on optimality of strategies: sampling designs and estimators, with numerous references, can be found, for example, in Cassel et al. (1977), and for a recent survey see Rinott (2008).

In Section 2 we provide all definitions and notations. In Section 3 we show that for our purposes randomized and behavioral estimators are equivalent. Thus we can choose either formulation of randomization according to our convenience. In Section 4 we describe the form of invariant estimators and some of their properties. We study best invariant-symmetric estimators under simple random sampling, and determine them explicitly in certain interesting cases. Sample quantiles, that is, quantiles of the empirical distribution function, provide a standard way of estimating the corresponding population quantiles. However, the estimators we propose and study in Section 4 are not always identical to the sample quantiles; also, that they may depend on the loss function under consideration. Furthermore, they may not be unique. In Section 5 we bring minimax results for invariant and non-invariant classes of estimators. Randomized estimators play a part in the proofs. Such estimators appear also when unbiasedness is desired. They are defined and studied in Malinovsky (2009).

2. Definitions and notations.   Most of the definitions and notations, with references, appear in Ferguson (1967), or Rinott (2008).

We consider a size N **finite population** of values of some measurement. Let $x = (x_1, x_2, ..., x_N)$ be the N-dimensional vector of population values, where x_j is a real number associated with the unit labeled $j \in \mathcal{N} := \{1, ..., N\}$, the label set.

We assume that $x \in \Upsilon$, a known **parameter space**. For simplicity we shall consider only parameter spaces of the type $\Upsilon = \{(x_1, x_2, ..., x_N) : x_i \in \mathbb{R}, \ x_i \text{ distinct}\}$, where $\mathbb{R}$ denotes the real line. Note that Υ is *symmetric* in the sense that if $x \in \Upsilon$ then so is any permutation of the coordinates of x. The assumption that the coordinates of x are distinct is not essential, but making it helps avoid various technicalities, and the same is true with regard to the assumption $x_i \in \mathbb{R}$, and we could assume that $x_i \in \Lambda$ where Λ is, for example, some known interval, finite or infinite. We will comment on such possibilities only briefly.

The **population distribution function** F_x is defined by

$$(2.1) \qquad F_x(t) = \frac{1}{N} \sum_{j=1}^{N} I_{(-\infty,\, t]}(x_j) = \frac{1}{N} \sum_{j=1}^{N} I_{[x_j,\, \infty)}(t).$$

F_x is an unknown **parameter** which is a function of the parameter x. Using the assumption that the coordinates of x are distinct, we can also write

$$F_x(t) = \frac{j}{N} \ \text{ for } x_{(j)} \le t < x_{(j+1)},\ j = 0, 1, ...N,\ (x_{(0)} := -\infty,\ x_{(N+1)} := \infty),$$

where $x_{(1)} < x_{(2)} < ... < x_{(N)}$ are the order statistics of x. In particular $F_x(x_{(j)}) = \frac{j}{N}$.

The k-th population **quantile** for a given $x \in \Upsilon$ is $\inf\{\theta \in \mathbb{R} : F_x(\theta) \ge k/N\}$.

Our goal is to estimate quantiles, where for a given estimate a of the k-th quantile, the **loss function** is of the type

$$(2.2) \qquad L(a, x) = G(|F_x(a) - \frac{k}{N}|), \quad a \in \mathbb{R},$$

for $k = 1, ..., N$, where G is a nonnegative increasing function. Some of our results focus on special cases of such G. Note that $|F_x(a) - \frac{k}{N}|$ vanishes if a is the k-th quantile, and otherwise it counts the deviation of a from the estimated quantile in terms of number of ordered population units by which they differ.

A parameter $\theta = \theta(x)$ is said to be ***symmetric*** if it remains constant under permutations of the coordinates of x. Clearly the examples given above, F_x and $x_{(k)}$ are symmetric parameters, and so is the population total $\theta(x) = \sum_{i=1}^{N} x_i$, and most of the common parameters of interest. Also, if for some θ, $F_x(\theta) \geq k/N$ for some x, then the same holds for any permutation of x since F_x is symmetric. Therefore the population quantiles are also symmetric.

A loss function $L(a, x)$ is said to be **symmetric** if it remains constant when x is replaced by any permutation of its coordinates for any a. It is clear that the loss (2.2) is symmetric since F_x is symmetric.

A **sampling design** $\mathcal{P}$ is a probability function on the space of all subsets S of $\mathcal{N}$. We assume noninformative sampling, that is, the probability $\mathcal{P}(S)$ does not depend on the parameter x. **Simple random sampling** *without replacement* of size n is denoted by $\mathcal{P}_s$ and satisfies $\mathcal{P}_s(S) = 1/\binom{N}{n}$ if $|S| = n$, and zero otherwise, where $|S|$ denotes the size of S.

The **data** consist of the set of pairs $\{(i, x_i) : i \in S\}$, that is , the x-values in the sample S and their corresponding labels. We set

$$(2.3) \qquad D = D[S, x] = \{(i, x_i) : i \in S\}.$$

We denote the set of all such D by $\mathcal{D}$. The notation $D[S, x]$ as defined above is sometimes convenient, however, is does not reflect the the pairing $(i, x_i) : i \in S$ which is part of the data, and the fact that the data depends on x only through x_i's such that $i \in S$.

By sufficiency arguments, Basu (1958) (also Cassel et al. (1977) and Rinott (2008)), the order in which the sample was drawn (if defined and known) and repetitions of units, if the sampling procedure allows it, provide no information. Since the relevant data consist only of the set of drawn labels S and their x-values, we shall only consider designs $\mathcal{P}$ on the space of unordered subsets of $\mathcal{N}$ with no repetitions.

We consider here only sampling designs having a fixed sample size, $|S| = n$, say; that is, the sample consists of n distinct units. Set $\mathbf{X} = \{x_i : i \in S\}$ and let $\mathbf{Y} = (Y_1, ..., Y_n) = (X_{(1)}, ..., X_{(n)})$ denote the ordered values (order statistics) of $\mathbf{X}$. We have $Y_1 < ... < Y_n$. In order to show the dependence on S, we often use the notation $x_S = \{x_i : i \in S\}$ instead of $\mathbf{X}$.

A (nonrandomized) **estimator** t is a real valued function $t(D)$ of the data. The space of such estimators is denoted by $\mathcal{T}$. We will also use the notation $t(S, x)$ and $t(\{(i, x_i) : i \in S\})$ for $t(D)$.

An estimator $t = t(D)$ is said to be **symmetric** if it depends only on the x-values in the sample, and not on their labels. In other words, the estimator

depends only on $\mathbf{X}$ and not on S. Thus, if $\{x_i : i \in S\} = \{x'_i : i \in S'\}$ for some $x, x' \in \Upsilon$ and samples S, S', then $t(D[S, x]) = t(D[S', x'])$. The class of all symmetric estimators is denoted by $\mathcal{T}_S$. It is trivial but important to note that without information on S, the information in $\mathbf{X} = \{x_i : i \in S\}$ is the same as in $\mathbf{Y}$. Hence for symmetric estimators we may write $t(\mathbf{X}) = t(\mathbf{Y})$, and also $t(x_S)$. The best known example of a non symmetric estimator is the Horvitz-Thompson estimator of the finite population total, $t_{HT}(D) := \sum_{i \in S} x_i / \alpha_i$, where the observation having label i is inversely weighted by the inclusion probability of the i-th unit according to the sampling design $\mathcal{P}$, $\alpha_i = P_{\mathcal{P}}(i \in S)$. On the other hand the simple sample mean, or the median and other sample quantiles, for example, are all symmetric.

A pair $(\mathcal{P}, t)$ consisting of a sampling design and an estimator is called a **strategy**.

The **risk** of a strategy $(\mathcal{P}, t)$ for a given $x \in \Upsilon$ is the expected loss defined by

$$(2.4) \qquad R(\mathcal{P}, t; x) = E_{\mathcal{P}} L(t(D), x) = \sum_S L(t(D[S, x]), x) \mathcal{P}(S).$$

For the next definition we need to consider the class of nonrandomized estimators $\mathcal{T}$ as a measure space. As in Ferguson (1967) we do not specify a sigma-field, however, we assume that singletons, that is, sets consisting of a single nonrandomized estimator, are in the sigma-field.

A probability distribution δ on the space of nonrandomized estimators $\mathcal{T}$, is called a **randomized estimator**. The space of all randomized estimators is denoted by $\mathcal{T}^*$. We define

$$(2.5) \qquad R(\mathcal{P}, \delta; x) = ER(\mathcal{P}, T; x) = \sum_S \int_{\mathcal{T}} L(t, x) d\delta(t) \mathcal{P}(S),$$

where T is a random variable taking values in $\mathcal{T}$, whose distribution is given by δ, and the integral with respect to $d\delta(t)$ has to be properly defined over the function space $\mathcal{T}$.

A randomized estimator is said to be **symmetric** if δ is concentrated on nonrandomized symmetric estimators. The class of such estimators is denoted be $\mathcal{T}_S^*$.

A **behavioral estimator** is defined by $\delta = \{\delta_D\} = \{\delta_{S,x_S}\}$, where for each possible data $D \in \mathcal{D}$, δ_D is a distribution on $\mathbb{R}$, with the interpretation that if D is observed, then a value in $\mathbb{R}$ is chosen according to δ_D as an estimate of the quantile in question. A behavioral estimator is said to be **symmetric** when the distributions δ_D depend only on x_S and not on the sampled labels.

For behavioral estimators, letting $Z \in \mathbb{R}$ be distributed according to δ_D we have

$$(2.6) \qquad R\left(\mathcal{P}, \delta; x\right) = ER\left(\mathcal{P}, Z; x\right) = \sum_S \int_{\mathbb{R}} L(z, x) \delta_{S, x_S}(dz) \mathcal{P}(S).$$

We remark that under the present setup the classes of behavioral and randomized rules are equivalent by WaldWolf (1951). In Section 3 we show that this equivalence holds also for invariant symmetric estimators, which are defined next. Therefore when discussing randomized estimators we consider either formulation according to our convenience and we use the same notation as defined above for randomized estimators, $\mathcal{T}^*$ and $\mathcal{T}_S^*$, also for the classes behavioral and symmetric behavioral estimators.

Given a function $\varphi : \mathbb{R} \to \mathbb{R}$, we extend its operation to vectors in the parameter space, by $\varphi(x) = (\varphi(x_1), ..., \varphi(x_N))$, to samples by $\varphi(\mathbf{Y}) = (\varphi(Y_1), ..., \varphi(Y_n))$, and to data by $\varphi(D) = \{(i, \varphi(x_i)) : i \in S\}$ and $\varphi(x_S) = \{\varphi(x_i)) : i \in S\}$. Let Φ denote the *group* of all *strictly increasing continuous* functions from $\mathbb{R}$ onto $\mathbb{R}$ (bijections).

In the case that in the parameter space we assume $x_i \in \Lambda$, an interval, say, then we assume that Φ consists of similar extensions of strictly increasing continuous functions from Λ onto Λ.

A nonrandomized estimator $t \in \mathcal{T}$ is said to be **invariant** if for all D and and all $\varphi \in \Phi$, we have

$$(2.7) \qquad\qquad\qquad t(\varphi(D)) = \varphi(t(D)).$$

The class of nonrandomized invariant estimators is denoted by $\mathcal{T}_I$, and the subclass of nonrandomized, invariant and symmetric estimators is denoted by $\mathcal{T}_{IS}$.

A randomized estimator $\delta \in \mathcal{T}^*$ is said to be **invariant** if δ, as a probability distribution over $\mathcal{T}$, assigns all its mass to the subset $\mathcal{T}_I$ of invariant nonrandomized estimators. The class of invariant randomized estimators is denoted by $\mathcal{T}_I^*$. A randomized estimator $\delta \in \mathcal{T}^*$ is said to be **invariant-symmetric** if δ, as a probability distribution over $\mathcal{T}$, assigns all its mass to the subset $\mathcal{T}_{IS}$ of invariant and symmetric nonrandomized estimators. The class of invariant-symmetric randomized estimators is denoted by $\mathcal{T}_{IS}^*$.

A behavioral estimator is said to be **invariant** if $Z_D \sim \delta_D$ satisfies $Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D)$ for all $D \in \mathcal{D}$, where $\overset{\mathcal{L}}{=}$ denotes equality of distributions (laws).

An **equalizer** estimator with respect to a design $\mathcal{P}$ is an estimator δ such that $R(\mathcal{P}, \delta; x) = C$ for some constant C, for all $x \in \Upsilon$.

Given a design $\mathcal{P}$, an estimator δ_1, is said to be **as good as** an estimator δ_2, if $R(\mathcal{P}, \delta_1; x) \leq R(\mathcal{P}, \delta_2; x)$ for all $x \in \Upsilon$. The estimator δ_1, is said to be **better** than an estimator δ_2, if $R(\mathcal{P}, \delta_1; x) \leq R(\mathcal{P}, \delta_2; x)$ for all $x \in \Upsilon$ and $R(\mathcal{P}, \delta_1; x) < R(\mathcal{P}, \delta_2; x)$ for at least one $x \in \Upsilon$, and **equivalent** to an estimator δ_2, if $R(\mathcal{P}, \delta_1; x) = R(\mathcal{P}, \delta_2; x)$ for all $x \in \Upsilon$. An estimator δ is said to be **admissible** if there exists no estimator better than δ.

An estimator having a property $\mathcal{C}$, that is as good as any other estimator having this property, is called a **best-$\mathcal{C}$** estimator. We shall consider $\mathcal{C} =$ the property of being invariant, or invariant and symmetric, or invariant and symmetric and unbiased (the latter are discussed only in Malinovsky (2009)).

3. Behavioral and randomized estimators. By a well known result in WaldWolf (1951), see also Ferguson (1967) and Kirschner (1976), behavioral and randomized estimators are equivalent in our problem. It may happen in certain situations that the classes of behavioral and randomized rules are equivalent, whereas, the classes of **_invariant_** behavioral and randomized rules are not equivalent. See, for example Ferguson (1967) p. 153. However, in our case the classes of symmetric invariant behavioral and invariant randomized rules are also equivalent. The result is close to that of Ferguson (1967) p. 197.

PROPOSITION 3.1. *For the group of transformations Φ defined above, the classes of symmetric invariant behavioral and symmetric invariant randomized estimators are equivalent.*

PROOF. The class of behavioral (invariant) estimators contains the class of randomized (invariant) estimators. For details see Ferguson (1967). We show that in our case the converse is also true, that is, given a symmetric invariant behavioral estimator $\delta = \{\delta_D\}$, we construct an equivalent symmetric invariant randomized estimator.

Consider a symmetric invariant behavioral estimator. Since it is symmetric we can write δ_{x_S} for δ_D. Let $Z_{x_S} \sim \delta_{x_S}$, and for simplicity of notation we now write χ for the set x_S. Choose χ_0, a particular point in the sample space, and a random variable $Z_{\chi_0} \sim \delta_{\chi_0}$. For each χ in the sample space choose $\varphi_\chi \in \Phi$ such that $\varphi_\chi(\chi_0) = \chi$. Define $\tilde{Z}_\chi = \varphi_\chi(Z_{\chi_0})$. This constructs a randomized estimator as follows: consider the nonrandomized function $t_a(\chi) = \varphi_\chi(a)$ for each $a \in \mathbb{R}$. Then $\tilde{Z}_\chi$ is distributed as the randomized estimator $t_a(\chi)$, with $a = Z_{\chi_0} \sim \delta_{\chi_0}$. Note that the invariance of the behavioral estimator Z_χ implies that $Z_\chi \overset{\mathcal{L}}{=} \varphi_\chi(Z_{\chi_0})$, and therefore

marginal distribution of $\tilde{Z}_\chi$ is the same as that of Z_χ, and therefore they are equivalent.

It remains to show that the constructed randomized estimator is invariant which means that the nonrandomized estimators $t_a(\chi)$ are invariant, that is, $t_a(\varphi(\chi)) = \varphi(t_a(\chi))$ with probability 1 with respect to $a \sim \delta_{\chi_0}$. This follows from $\varphi(t_a(\chi)) = \varphi(\varphi_\chi(Z_{\chi_0})) \stackrel{\mathcal{L}}{=} Z_{\varphi(\varphi_\chi(\chi_0))} = Z_{\varphi(\chi)} = Z_{\varphi_{\varphi(\chi)}(\chi_0)} \stackrel{\mathcal{L}}{=} \varphi_{\varphi(\chi)}(Z_{\chi_0}) = t_a(\varphi(\chi))$. In particular we have $\varphi(\varphi_\chi(Z_{\chi_0})) \stackrel{\mathcal{L}}{=} \varphi_{\varphi(\chi)}(Z_{\chi_0})$. By Lemma 3.1 below this implies the equality almost surely, and then by the above relations $\varphi(t_a(\chi)) = t_a(\varphi(\chi))$ almost surely, and the proof is complete. □

LEMMA 3.1.   *If V is a random variable and f and g are strictly increasing continuous functions such that $g(V) \stackrel{\mathcal{L}}{=} h(V)$ then $g(V) = h(V)$ almost surely.*

PROOF. It suffices to prove that if $g(V) \stackrel{\mathcal{L}}{=} V$ then $g(V) = V$ with probability one. Let F denote the distribution function of V, and denote g^{-1} by h. Then the assumed equality in distribution is equivalent to $F(h(v)) = F(v)$ for all v in the support of F.

If F is strictly increasing then the assumption becomes $F(h(v)) = F(v)$ for all v, which implies $h(v) = v$ for all v. If F is not strictly increasing then almost the same argument works for points of increase of F, whereas other points have F probability zero. More specifically, if v is in the support of F then either $F(v + \varepsilon) > F(v)$ for all sufficiently small ε, or $F(v - \varepsilon) < F(v)$ for all sufficiently small ε. In the first case, for example, we cannot have $F(h(v)) = F(v)$ for any h satisfying $h(v) > v$. If $h(v) < v$, then for some ε we have $h(v+\varepsilon) < v$ by continuity of h. Then $F(h(v+\varepsilon)) < F(v) < F(v+\varepsilon)$ contradicting the assumption that $F(h(v)) = F(v)$ for all v in the support of F. □

4. Invariant estimators.

4.1. *General form of invariant and symmetric estimators.* Proposition 4.1 below in an infinite population setup, appears in Uhlmann (1963), Ferguson (1967), p.153, Ex 4.2.3, and Zieliński (1999). They show that the only non-randomized invariant estimators are of the form Y_j, for some j independent of the data. In the above references the data consist of a sample of iid observations from an unknown distribution or equivalently from an infinite population. In our case of finite population the data include also the labels of the observations. In the proposition below we show that invariant estimators are of the form $Y_{j(D)}$. However, the fact that the data now contain

labels makes it impossible to conclude that j is independent of the data as in the infinite population case where labels do not exist. If symmetry is also assumed, then by definition $j(D)$ no longer depends on S, and then it can be shown that it does not depend on D at all. Further subtle issues that arise in the presence of labels appear in Theorem 5.1 and the lemmas around it.

The next proposition is stated and proved for behavioral estimators, hence it holds also for randomized estimators, including the first part where symmetry is not assumed and therefore Proposition 3.1 does not apply.

PROPOSITION 4.1.   *The behavioral invariant estimators are of the form* $t(D) = Y_{J(D)}$, *where* $J(D)$ *is a random variable taking values in* $\{1, ..., n\}$, *having distribution that depends on* D. *Moreover, the distribution of* $J(D) = J(D[S, x])$ *depends only on* S *and not on* x.

Also, symmetric behavioral invariant estimators are of the form $t(D) = t(\mathbf{Y}) = Y_J$, *where* J *is a random variable taking values in* $\{1, ..., n\}$, *having distribution that is not a function of the data.*

PROOF. The statement of this proposition is an adaptation of exercise 3, p. 197 in Ferguson (1967), which deals with iid sampling from continuous distributions. The following proof is a simplification of the proof in Ferguson's web site.

A behavioral estimator is defined by a collection of random variables $Z_D \sim \delta_D$, taking values in the decision space, which in our case is $\mathbb{R}$ or a subset thereof. Invariance means $Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D)$. In particular this holds for all strictly increasing continuous functions φ, which leave $X_1, ..., X_n$ fixed. The set of such functions φ is denoted by Φ'.

For $\varphi \in \Phi'$ we have

$$(4.1) \qquad Z_D = Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D).$$

It follows that the support of Z_D must be contained in the set $\{Y_1, \ldots, Y_n\}$; otherwise, it is easy to construct $\varphi \in \Phi'$ such that $\text{Support}[Z_D] \neq \text{Support}[\varphi(Z_D)]$, contradicting (4.1). Therefore, behavioral invariant estimators are of the form $Y_{J(D)}$, where $J(D)$ is a random variable taking values in $\{1, ..., n\}$, whose distribution may depend on D.

Moreover, the above representation implies that $Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Y_{J(\varphi(D))})$ and invariance means $Z_{\varphi(D)} \overset{\mathcal{L}}{=} \varphi(Z_D) = \varphi(Y_{J(D)})$. It follows that $\varphi(Y_{J(\varphi(D))}) \overset{\mathcal{L}}{=} \varphi(Y_{J(D)})$, and therefore

$$(4.2) \qquad J(\varphi(D)) \overset{\mathcal{L}}{=} J(D).$$

Let $D^{'} = \{(i, x_i^{'}) : i \in S\}$. For any vector $x = (x_1, ..., x_N)$ there exists $\varphi \in \Phi$, such that $\{\varphi(x_i^{'}) : i \in S\} = \{x_i : i \in S\}$. This with (4.2) implies that $J(D)$ may depend only on S.

The last part of the Proposition follows readily since by symmetry $J(D)$ does not depend on S either, and therefore it does not depend on D.   □

COROLLARY 4.1.   *There are only n nonrandomized symmetric invariant estimators, of the form*

(4.3) $$t(D) = Y_j, \; j = 1, ..., n.$$

REMARK 4.1.   *In the case that the parameter space is*

$$\Upsilon = \{(x_1, x_2, ..., x_N) : x_i \in \Lambda, \; x_i \; distinct\},$$

where $\Lambda = [a, b]$, say, then the estimators $Y_0 \equiv a$ and $Y_{n+1} \equiv b$ are also invariant and must be taken into account. In this case the above results hold, but the range of J or j changes from $\{1, ..., n\}$ to $\{0, ..., n+1\}$.

4.2. *Best invariant-symmetric estimators under simple random sampling.* In this subsection we consider only simple random sampling, that is, $\mathcal{P}_s$. We now deal with symmetric estimators. In Section 5 we shall consider nonsymmetric estimators.

For the estimators of (4.3) we have

LEMMA 4.1.   *Under $\mathcal{P}_s$ and the loss (2.2) any estimator Y_j is an equalizer.*

PROOF. It is easy to see that the distribution of $NF_x(Y_j)$ under $\mathcal{P}_s$ is the same as the distribution of the j-th order statistic in a simple random sample of size n from $\{1, \ldots, N\}$. Clearly, this distribution does not depend on the parameter x, and the result follows. More explicitly, the distribution of $F_x(Y_j)$ under $\mathcal{P}_s$ for $j = 1, \ldots, n$ is

$$P_{\mathcal{P}_s}\left(F_x(Y_j) = \frac{m}{N}\right) = \frac{\binom{m-1}{j-1}\binom{N-m}{n-j}}{\binom{N}{n}}; \; m = j, \ldots, N-n+j.$$

See, e.g., Wilks (1962), p.243, Arnold et al. (1992) p.54, David and Nagaraja (2003) p.23. It follows that the distribution of $F_x(Y_j)$ under $\mathcal{P}_s$ does not depend on the parameter $x \in \Upsilon$ and under (2.2) the estimator Y_j is therefore an equalizer.   □

We remark that the estimator $Y_0 \equiv a$ described in Remark 4.1 is not an equalizer, while $Y_{n+1} \equiv b$ is an equalizer.

Lemma 4.1 follows from the easy fact that under $\mathcal{P}_s$ the distribution of $F_x(Y_j)$ is independent of x. However, Lemma 4.1 does not hold in general, that is, for any sampling design. For example, if $n = 2, N = 3, k = 1$, and $t(D) = Y_1$, then a design that chooses $S = \{1, 2\}$ with probability $= 1$ has the risk (and loss) $G(0)$ under (2.2) if $x_1 < x_2 < x_3$. However, if $x_3 < x_2 < x_1$, the risk is $G(|\frac{1}{3}|)$.

DEFINITION 4.1.   *Define*
(4.4)
$$j^* := j^*_{G,k} = \arg\min_j R(\mathcal{P}_s, Y_j; x) = \arg\min_{j \in \{1, \ldots, n\}} E_{\mathcal{P}_s} G\left(\left| F_N(Y_j) - \frac{k}{N} \right|\right).$$

If the minimum is not unique, then one can view $j^*_{G,k}$ as the set where the minimum obtains, or one of the minimizers.

REMARK 4.2.   *Below we discuss the estimator Y_{j^*}. Proposition 4.3 gives an explicit expression for j^* for square error loss, that is, $G(u) = u^2$. For example, when $N = 100, n = 10$, and $k = 79$, one gets $j^* = 9$. Note, however, that $(j^* - 1)/n = 8/10 > k/N = 79/100$, so that here $Y_{j^*} = Y_9$ is clearly not the sample quantile corresponding to the k-th population quantile; this sample quantile is at most Y_8. Thus our estimators are not always the "natural" sample quantiles, although in general they are close. One can define such quantiles as any $Y_{\bar{j}}$ such that $\frac{k}{N} - \frac{1}{n} < \frac{\bar{j}}{n} < \frac{k}{N} + \frac{1}{n}$. The above example shows that j^* does not always satisfy the latter inequalities, although it does very often.*

In general, j^*, may depend on G. For example, when $L(a, x) = |F_x(a) - \frac{k}{N}|^r$, that is, when $G(u) = u^r$, j^* depends on r. Consider $N = 9$, $n = 7$, $k = 2$; direct calculations show that for $r \leq c$ we have $j^* = 2$ whereas $r > c$ implies $j^* = 1$, where $c = \log(17/3)/\log(2) \approx 2.5$. This means that the above estimator of the second population quantile, $t = Y_{j^*}$ depends on the loss function. This is a natural but somewhat undesirable state of affairs, since statisticians often do not have a precise loss function in mind.

By definition and Proposition 4.1, we obtain

COROLLARY 4.2.   *Under $\mathcal{P}_s$, Y_{j^*} is the best nonrandomized invariant-symmetric estimator, that is, it is best in the class of nonrandomized, symmetric and invariant estimators.*

A stronger result holds:

THEOREM 4.1.   *Under $\mathcal{P}_s$, the (nonrandomized) estimator Y_{j^*} is also the best invariant-symmetric estimator in $\mathcal{T}_{IS}^*$, that is, the estimator minimizing*

$$(4.5) \qquad E_{\mathcal{P}_s} EG(|F_N(t(D)) - \frac{k}{N}|)$$

among randomized and behavioral invariant-symmetric estimators $t(D)$.

   Note that the expectation $E_{\mathcal{P}_s}$ in (4.5) is with respect to simple random sampling, while the the second expectation is with respect to the randomness of $t(D)$.

   PROOF. According to Lemma 4.1 the estimator $t(D) = Y_j$ is an equalizer. From Corollary 4.1 it follows that every nonrandomized symmetric invariant estimator is of the form Y_j for some j and the best invariant-symmetric among nonrandomized estimators is Y_{j^*}, and therefore

$$(4.6) \quad E_{\mathcal{P}_s} G(|F_N(Y_{j^*}) - \frac{k}{N}|) \le \sum_{j=1}^{n} \alpha_j E_{\mathcal{P}_s} G(|F_N(Y_j) - \frac{k}{N}|) \ \ \forall x,$$

   for any $\alpha_1, ..., \alpha_n$ such that $\alpha_i \ge 0 \ \forall i = 1, ..., n$ and $\sum_{i=1}^{n} \alpha_i = 1$.

Note that any risk of a randomized invariant-symmetric estimator can be represented by the right-hand side of (4.6). Together with Proposition 4.1, it follows that the estimator Y_{j^*} is the best among randomized invariant-symmetric estimators in $\mathcal{T}_{IS}^*$, and by Proposition 3.1 it is also best among behavioral invariant-symmetric estimators.   □

   We next describe two important cases where j^* is known, and in these cases, by Theorem 4.1, the best randomized of behavioral invariant-symmetric estimator is given explicitly. First, for estimating the median when N and n are odd, that is, $k = \frac{N+1}{2}$, we have $j^* = \frac{n+1}{2}$ for any (increasing) G; see Theorem 4.2 below. The other case is when $G = u^2$, given in Proposition 4.3.

   The following theorem is a special case of a result in Malinovsky and Rinott (2009). The result seems obvious, but the proof requires more calculations than expected; it has a simple combinatorial flavor since under simple random sampling the quantities $NF_x(Y_j)$ appearing below are distributed as the order statistics of a simple random sample of size n from $\{1, \ldots, N\}$. Below $\ge_{st}$ stands for "stochastically larger".

THEOREM 4.2.   *Let $Y_1, \ldots, Y_n$ be the order statistics of a simple random sample without replacement from a finite population consisting of N distinct values, where n and N are odd. Then*

$$(4.7) \qquad \left| F_x(Y_j) - \frac{N+1}{2N} \right| \geq_{st} \left| F_x(Y_{\frac{n+1}{2}}) - \frac{N+1}{2N} \right| \quad for \ \ j = 1, \ldots, n.$$

It follows that in this case $j^* = \frac{n+1}{2}$.

If the parameter space is such that $x_i \in [a, b]$, then it is easily seen that Theorem 4.2 holds also for $j = 0$ and $j = n+1$, where $Y_0 \equiv a$ and $Y_{n+1} \equiv b$.

Next we compute $j^* = j^*_{G,k}$ explicitly $G(u) = u^2$, that is, for the case of square error, where $L(a, x) = |F_x(a) - \frac{k}{N}|^2$. The following lemma is useful for this purpose. It can be found in Wilks (1962), p. 244.

LEMMA 4.2.

$$E_{\mathcal{P}_s}\left(F_x(Y_j)\right) = \frac{j}{N}\frac{N+1}{n+1}, \ \ j = 1, \ldots, n$$

$$E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) = \frac{j}{N}\frac{N+1}{n+1}\left(\frac{(j+1)(N+2)}{N(n+2)} - \frac{1}{N}\right), \ \ j = 1, \ldots, n.$$

We remark that for estimating the median, for example, in the case of odd N and n, we have "unbiasedness" in the sense that Theorem 4.2 implies $j^* = \frac{n+1}{2}$, and by the first equality in Lemma 4.2 $E_{\mathcal{P}_s}\left(F_x(Y_{j^*})\right) = \frac{N+1}{2N}$. In general, such unbiasedness may require randomized estimators.

THEOREM 4.3. $j^* = \arg\min_{j \in \{1,\ldots,n\}} E_{\mathcal{P}_s}(|F_N(Y_j) - \frac{k}{N}|)^2$ *is the value* $j^* \in \{1, \ldots, n\}$ *that is the nearest integer to* $j^{**} = \frac{n+2}{N+2}\left(k + \frac{1}{2}\right) - \frac{1}{2}$.

PROOF. From

$$E_{\mathcal{P}_s}\left(F_x(Y_j) - \frac{k}{N}\right)^2 = E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right) + \frac{k^2}{N^2}$$

it follows that

$$\arg\min_{1 \leq j \leq n}\left\{E_{\mathcal{P}_s}\left(F_x(Y_j) - \frac{k}{N}\right)^2\right\} = \arg\min_{1 \leq j \leq n}\left\{E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right)\right\}.$$

Using Lemma 4.2 we have:

$$
E_{\mathcal{P}_s}\left(F_x^2(Y_j)\right) - 2\frac{k}{N}E_{\mathcal{P}_s}\left(F_x(Y_j)\right)
$$
$$
= \frac{j}{N}\frac{N+1}{n+1}\left(\frac{(j+1)(N+2)}{N(n+2)} - \frac{1}{N}\right) - 2\frac{k}{N}\frac{j}{N}\frac{N+1}{n+1}
$$
$$
= \frac{(N+1)(N+2)}{N^2(n+1)(n+2)}j^2 + \left(\frac{(N+1)(N+2)}{N^2(n+1)(n+2)} - \frac{N+1}{N^2(n+1)}(2k+1)\right)j = f(j).
$$

The last expression f(j) is a convex parabola as a function of the continuous variable j which attains a minimum at the point $= \frac{n+2}{N+2}\left(k+\frac{1}{2}\right) - \frac{1}{2}$. This point is not necessarily an integer.

Setting $j^* = \arg\min_{j\in\{1,\dots,n\}} E_{\mathcal{P}_S}\left(F_x(Y_j) - \frac{k}{N}\right)^2$ it is clear by symmetry of the parabola $f(j)$ around its minimum, that j^* is the nearest integer to the minimum point of f. $\qquad\square$

REMARK 4.3.   *If N and n are odd and we estimate the median $\left(k = \frac{N+1}{2}\right)$, then, $j^* = j^{**} = \frac{n+1}{2}$ (the sample median).*
   The estimator defined in Theorem 4.3 is not always unique: if $N = 7, n = 4, k = 4$, then $\frac{n+2}{N+2}\left(k+\frac{1}{2}\right) - \frac{1}{2} = 2.5$. Hence the estimators for the $\frac{4}{7}$ quantile (Median) are both Y_2 or Y_3, as well as any estimator which randomizes between these two estimators.
   If $N = 20, n = 6, k = 5$, then $\frac{n+2}{N+2}\left(k+\frac{1}{2}\right) - \frac{1}{2} = 1.5$. Hence for the $\frac{4}{20}$ quantile the estimators of Theorem 4.3 are Y_1 or Y_2.

COROLLARY 4.3.   *The estimator given in Theorem 4.3 is unique if either n is odd or both N and $\dfrac{n}{2}$ are even.*

PROOF. In the proof of Proposition 4.3 it was shown that $E_{\mathcal{P}_s}\left(F_x(X_{(j)}) - \frac{k}{N}\right)^2$ is a convex and symmetric function with minimum at the point $j^{**} = \frac{n+2}{N+2}\frac{2k+1}{2} - \frac{1}{2}$. Hence, is clear that if the number $\frac{n+2}{N+2}\frac{2k+1}{2}$ is not an integer then the estimator t^* is unique. If n is odd, then the numerator of the above ratio is odd, while the denominator is clearly even. If N and $\frac{n}{2}$ are even, then $\frac{n+2}{N+2}\frac{2k+1}{2} = \frac{\left(\frac{n}{2}+1\right)(2k+1)}{N+2}$, and again the numerator is odd while the denominator is even. $\qquad\square$

Clearly, there exist many other cases where $\frac{n+2}{N+2}\frac{2k+1}{2}$ is not an integer not covered above.

REMARK 4.4. *The uniqueness of Corollary 4.3 does not hold for absolute loss. For example, when $N = 7$ and $n = 3$ and our purpose is to estimate the quntile $\frac{5}{7}$, the best invariant-symmetric estimator under absolute loss is Y_2 or Y_3.*

5. Minimax results for non symmetric or non invariant estimators.

5.1. *Minimax strategies and invariance.* In this section we study minimax strategies with invariant estimators. In particular we show that in minimax strategies the estimators are necessarily symmetric. Symmetrization of strategies as (5.3) below appears in Blackwell and Girshick (1954), and in Kiefer (1957) with references to work of Hunt and Stein from the 1940s. The required notations and formulations explained next follow Stenger (1979) and Rinott (2008), where further references can be found.

Let π be a permutation of $\mathcal{N}$. For $S \subseteq \mathcal{N}$ we define $\pi S = \{\pi i : i \in S\}$. For $x \in \Upsilon$ let πx be the parameter vector having coordinates

$$(5.1) \qquad (\pi x)_i = x_{\pi^{-1}i}.$$

Thus, the group Π of permutations of $\{1, 2, ..., N\}$ can also be seen as a group operating on the (symmetric) parameter space Υ, where the group operation is permutation of the coordinates.

Recall that an estimator (nonrandom, randomized or behavioral) is symmetric if $t(S, x)$ is only a function of x_S, that is, if $x_S = x'_{S'}$ for two sets of size n in $\mathcal{N}$, then $t(S, x) = t(S', x')$, where for a randomized estimator the latter equality stands for equality in distribution.

Given an estimator t, let $t_\pi(S, x) = t\left(\{(\pi i, x_i) : i \in S\}\right)$.

Given a strategy $(\mathcal{P}, t)$ with a fixed sample size and a nonrandomized estimator t, let t^* be the randomized estimator

$$(5.2) \qquad t^*(D[S, x]) = t_\pi(S, x) \text{ with probability } c\mathcal{P}(\pi S) \text{ for } \pi \in \Pi,$$

and for a randomized behavioral estimator $Z_D \sim \delta_D$ let t^* be the randomized behavioral estimator

$$(5.3) \qquad t^*(D[S, x]) = Z_{\{(\pi i, x_i): i \in S\}} \text{ with probability } c\mathcal{P}(\pi S) \text{ for } \pi \in \Pi,$$

where $c = \dfrac{1}{n!(N-n)!} = \dfrac{1}{N!\mathcal{P}_s(S)}$ is such that $\sum_{\pi \in \Pi} c\mathcal{P}(\pi S) = 1$, and $Z_{\{(\pi i, x_i): i \in S\}}$, $\pi \in \Pi$, are taken to be independent. Set $S = \{s_1, \ldots, s_n\}$. An equivalent formulation is

$$(5.4) \qquad t^*(\{(i, x_i) : i \in S\}) = Z_{(\{(\ell_i, x_{s_i}): i=1, \ldots, n\})} \text{ w. p. } c\mathcal{P}(\{\ell_1, \ldots, \ell_n\}),$$

for all $(\ell_1, \ldots, \ell_n)$ having distinct coordinates in $\mathcal{N}$.

Note that the probabilities $\frac{\mathcal{P}(\pi S)}{N! \mathcal{P}_s(S)}$ in (5.3) seem to depend on S, making t^* appear like a non symmetric estimator. However, from (5.4) we see that t^* is symmetric and depends only on x_S. Thus we have

LEMMA 5.1. t^* *is a symmetric (randomized) estimator.*

LEMMA 5.2.   *If t is invariant, then t^* is invariant.*

PROOF. By definition $t^*(D[S, \varphi(x)]) = Z_{(\pi S, \varphi(x))}$ with probability $c\mathcal{P}(\pi S)$. Since $Z_{(\pi S, \varphi(x))} \overset{\mathcal{L}}{=} \varphi(Z_{(\pi S, x)})$, invariance follows.   □

EXAMPLE 5.1.   *Consider $N = 3, n = 2$, and the sampling design $P(S_i) = q_i, i = 1, 2, 3$, $q_1 + q_2 + q_3 = 1$, where $S_1 = \{1, 2\}, S_2 = \{1, 3\}, S_3 = \{2, 3\}$. Consider the nonrandomized invariant nonsymmetric estimator*

$$t = \begin{cases} Y_1, & \text{if } 1 \in S \\ Y_2, & \text{if } 1 \notin S, \end{cases}$$

where, as defined in Section 2, Y_i are the sample order statistics. The corresponding estimator t^* of (5.3) is the symmetric (depending only on Y_1, Y_2 and independent of S) randomized estimator

$$t^* = \begin{cases} Y_1, & \text{w.p. } q_1 + q_2 \\ Y_2, & \text{w.p. } q_3 \ . \end{cases}$$

By Lemmas 5.1 and 5.2 and Proposition 4.1 we now know the form of t^* as follows:

THEOREM 5.1.   *If $t(D)$ is invariant, then the corresponding estimator t^* of (5.3) is a randomized estimator of the form $t^*(D) = Y_J$, where J is a random variable whose distribution is independent of the data D.*

EXAMPLE 5.2.   *Let $N = 3, n = 2$. Given a sample S of two elements, define $\ell = \ell(S) = \min\{i : i \in S\}$, and $m = m(S) = \max\{i : i \in S\}$. Consider the following nonrandomized invariant nonsymmetric estimator*

$$t = \begin{cases} Y_1, & \text{if } x_\ell < x_m \\ Y_2, & \text{if } x_\ell > x_m \ . \end{cases}$$

Then for $l_\pi := \min\{\pi i : i \in S\}$ *and* $m_\pi := \max\{\pi i : i \in S\}$

$$t_\pi = \begin{cases} Y_1, & \text{if } x_{\ell_\pi} < x_{m_\pi} \\ Y_2, & \text{if } x_{\ell_\pi} > x_{m_\pi} \ . \end{cases}$$

The corresponding estimator t^* *of* (5.2) *is the randomized symmetric estimator:*

$$\begin{cases} Y_1, & \text{with probability } \frac{1}{2} \\ Y_2, & \text{with probability } \frac{1}{2} \ . \end{cases}$$

A version of the next proposition appears with references as Proposition 13 in Rinott (2008). It is relevant in reducing considerations of minimax strategies to $\mathcal{P}_s$ and symmetric estimators.

PROPOSITION 5.1. *Let* $L(t, x)$ *be a symmetric loss function and as always let* Υ *be a symmetric parameter space. Given a strategy* $(\mathcal{P}, t)$ *with fixed sample size* n *and a behavioral (or randomized, or nonrandomized) estimator* t, *let* $t^*(D[S, x])$ *be the estimator defined by* (5.3). *Then*

$$(5.5) \qquad \sup_{x \in \Upsilon} R(\mathcal{P}_s, t^*; x) \leq \sup_{x \in \Upsilon} R(\mathcal{P}, t; x).$$

PROOF. Consider a behavioral estimator $Z_D \sim \delta_D$, and observe that in the present case (2.6) can be expressed as in the first equality below:

$$R(\mathcal{P}_s, t^*; x) = \sum_S \sum_\pi \frac{\mathcal{P}(\pi S)}{N! \mathcal{P}_s(S)} EL\left(Z_{\{(\pi i, x_i) : i \in S\}}, x\right) \mathcal{P}_s(S)$$

$$= \frac{1}{N!} \sum_S \sum_\pi \mathcal{P}(\pi S) EL\left(Z_{\{(\pi i, x_i) : i \in S\}}, x\right)$$

$$\stackrel{(1)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, x_{\pi^{-1} i}) : i \in S\}}, x\right) \stackrel{(2)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, x_{\pi^{-1} i}) : i \in S\}}, \pi x\right)$$

$$\stackrel{(3)}{=} \frac{1}{N!} \sum_\pi \sum_S \mathcal{P}(S) EL\left(Z_{\{(i, (\pi x)_i) : i \in S\}}, \pi x\right) = \frac{1}{N!} \sum_\pi R(\mathcal{P}, t; \pi x) \leq \sup_\pi R(\mathcal{P}, t; \pi x),$$

where the equality (1) follows by substituting S for πS, (2) by symmetry of L, and (3) by (5.1). Taking sup over $x \in \Upsilon$ yields the result. $\square$

THEOREM 5.2. *The strategy* $(\mathcal{P}_s, Y_{j^*})$, *with* j^* *defined in* (4.4) *is minimax among all strategies* $(\mathcal{P}, t)$ *consisting of a sampling design* $\mathcal{P}$ *having a*

fixed sample size n, and any randomized or behavioral invariant estimator $t(D)$, that is,

$$(5.6) \quad \inf_{(t \in T_I^*, \mathcal{P})} \sup_{x \in \Upsilon} E_{\mathcal{P}} EG(|F_x(t(D)) - \frac{k}{N}|) = \sup_{x \in \Upsilon} E_{\mathcal{P}_s} G(|F_x(Y_{j^*}) - \frac{k}{N}|),$$

where $E_{\mathcal{P}}$ stands for expectation with respect to the design $\mathcal{P}$ and E on the left-hand side is with respect to the randomness of t. Equivalently,

$$(5.7) \quad \inf_{(t \in T_I^*, \mathcal{P})} \sup_{x \in \Upsilon} R(\mathcal{P}, t(D); x) \geq \sup_{x \in \Upsilon} R(\mathcal{P}_s, Y_{j*}; x).$$

PROOF. By the first part of Proposition 4.1, we can restrict attention to estimators of the form $Y_{J(D)}$. Using Theorem 5.1 together with Proposition 5.1 we have

$$(5.8) \quad \sup_{x \in \Upsilon} E_{\mathcal{P}} EG(|F_x(Y_{J(D)}) - \frac{k}{N}|) \geq \sup_{x \in \Upsilon} E_{\mathcal{P}_S} EG(|F_x(t^*(D)) - \frac{k}{N}|),$$

where $t^*(D) = Y_J$ is the randomized estimator obtained from $t(D) = Y_{J(D)}$ by (5.3), and the distribution of J is independent of the data D.

Because, Y_J is a symmetric estimator we have from Theorem 4.1 for j^* defined in (4.4)

$$(5.9) \quad E_{\mathcal{P}_S} EG(|F_x(t^*(D)) - \frac{k}{N}|) \geq E_{\mathcal{P}_S} G(|F_x(Y_{j^*}) - \frac{k}{N}|).$$

Combining (5.8) and (5.9), we end the proof.   □

The next corollary follows from Theorems 4.2 and 5.2.

COROLLARY 5.1.   *For odd N and n, the strategy $\left(\mathcal{P}_s, Y_{\frac{n+1}{2}}\right)$ is minimax among all strategies $(\mathcal{P}, t)$ consisting of a sampling design $\mathcal{P}$ having a fixed sample size n, and a randomized or behavioral invariant estimator t, that is,*
$$(5.10)$$
$$\inf_{(t \in T_I^*, \mathcal{P})} \sup_{x \in \Upsilon} E_{\mathcal{P}} EG(|F_x(t(D)) - \frac{N+1}{2N}|) = \sup_{x \in \Upsilon} E_{\mathcal{P}_s} G(|F_x(Y_{\frac{n+1}{2}}) - \frac{N+1}{2N}|).$$

5.2. *Minimax results without invariance.* In this section we prove two results which compare the minimax risk of our estimators to classes of estimators which are not invariant. In Theorem 5.3 we focus for simlicity on the sample median, and compare it to non-invariant estimators whose distance from the median is bounded. In Theorem 5.4 we compare our quantile estimates to linear estimators.

An admissible equalizer estimator is minimax. In fact a somewhat weaker property suffices:

LEMMA 5.3.   *An equalizer estimator t_0 is minimax relative to some class of estimators if for any estimator t in the class, and any $\varepsilon > 0$ there exists $x \in \Upsilon$ such that*

$$(5.11) \qquad R(\mathcal{P}_s, t(x_S); x) > R(\mathcal{P}_s, t_0(x_S); x) - \varepsilon.$$

PROOF. If t_0 is not minimax, then for some t, $\sup_x R(\mathcal{P}_s, t(x_S); x) < \sup_x R(\mathcal{P}_s, t_0(x_S); x)$. Since t_0 is an equalizer it follows that $\sup_x R(\mathcal{P}_s, t(x_S); x) < R(\mathcal{P}_s, t_0(x_S); x)$ and therefore for some $\varepsilon > 0$ $\sup_x R(\mathcal{P}_s, t(x_S); x) < R(\mathcal{P}_s, t_0(x_S); x) - \varepsilon$, contradicting (5.11). □

Let N and n be odd, and set $t_0(x_S) = Y_{\frac{n+1}{2}}$. By Lemma 4.1 t_0 is an equalizer. The following theorem is of interest because it reflects the combinatorial nature of our structure.

THEOREM 5.3.   *Assume the loss function is of the form $L(a, x) = G(|F(a) - \frac{N+1}{2N}|)$, where G is convex and increasing. As usual, $\Upsilon = \{(x_1, x_2, ..., x_N) : x_i \in \mathbb{R}, \ x_i \ distinct\}$ and let the sampling design be $\mathcal{P}_s$. Then t_0 is minimax among nonrandomized symmetric estimators t satisfying*

$$(5.12) \qquad |t(x_S) - Y_{\frac{n+1}{2}}| < B \quad for \ some \ B;$$

that is, $\sup_x R(\mathcal{P}_s, t(x_S); x) \geq R(\mathcal{P}_s, t_0(x_S); x) \quad \forall x \in \Upsilon.$

The above result can be easily extended to randomized estimator which are a finite or countable mixture of nonrandomized estimators. More generally, recall that for a randomized estimator the risk R is defined by $R(\mathcal{P}, \delta; x) = \sum_S \int L(t, x) d\delta(t) \mathcal{P}(S)$, where the integral is with respect to the measure δ on the function space $\mathcal{T}$. Without going into measurability issues, any integral $\int L(t, x) d\delta(t)$ would be defined by finite sum approximations of the type $\sum_{t_i} L(t_i, x) \Delta(t_i)$ and therefore for the Theorem 5.3 applies also to randomized estimators.

Furthermore, given a strategy $(\mathcal{P}, t)$ we can use Proposition 5.1 to replace it by the strategy $(\mathcal{P}_s, t^*)$ with t^* defined in (5.2), and conclude that Theorem 5.3 holds without the assumption of symmetry on t, and for any sampling design $\mathcal{P}$, that is, $(\mathcal{P}_s, t_0)$ is minimax among all strategies $(\mathcal{P}, t)$ such that $\mathcal{P}$ has sample size n, and t is a nonrandomized estimator satisfying (5.12).

The latter condition may seem artificial: it does not hold for the sample mean, for example. However, since $Y_{\frac{n+1}{2}}$ is the most natural estimator of the population median, one may interpret condition (5.12) as a reasonable restriction suggesting that if an estimate is too far from the sample median, it should be corrected (or trimmed).

PROOF OF THEOREM 5.3. To prove (5.11) (with $\varepsilon = 0$) form an infinite set Γ of points in $\mathbb{R}$ such that each pair of points in Γ is spaced by more than B. Every set x_S of n data points in Γ satisfies either (a) $t(x_S) < Y_{\frac{n+1}{2}}$, or (b) $t(x_S) \geq Y_{\frac{n+1}{2}}$, where as before $Y_{\frac{n+1}{2}}$ is the median of x_S. By the infinite Ramsey Theorem, see, e,g, Graham et al (1990) page 19 Theorem A, there exists an infinite subset Δ of Γ such that either all its n-subsets x_S satisfy (a) above, or all satisfy (b). In the latter case, we take N point in Δ and form x, to obtain $R(\mathcal{P}_s, t(x_S); x) = R(\mathcal{P}_s, t_0; x)$ (here we use the B spacing).

It remains to consider the case that for the above x, all n-subsets x_S satisfy (a). Divide (partition) the set of $\binom{N}{n}$ possible samples into two subsets, A_1 and A_2, as follows: $A_1 = \{S : Y_{\frac{n+1}{2}} \leq x_{(N+1)/2}\}$, and $A_2 = \{S : Y_{\frac{n+1}{2}} > x_{(N+1)/2}\}$. Assume that the components of x are arranged in increasing order. For each $S = \{s_1, \ldots, s_n\}$ in A_2, its reflection around the median, $S' = \{N + 1 - s_1, \ldots, N + 1 - s_n\} \in A_1$, and $|F_x(t_0(x_S)) - \frac{N+1}{2N}|$ has the same value for S and S'. For any $S \in A_2$ there corresponds one point in A_1, it reflection. In fact $|A_2| < |A_1|$ since some point in A_1 have reflection also in A_1. Moreover, for $S \in A_2$ we have, due to condition (a), $|F_x(t_0(x_S)) - \frac{N+1}{2N}| = |F_x(t(x_S)) - \frac{N+1}{2N}| + 1/N$, and for $S \in A_1$ we have $|F_x(t_0(x_S)) - \frac{N+1}{2N}| = |F_x(t(x_S)) - \frac{N+1}{2N}| - 1/N$, where again the B spacing was used. It follows that

$$R(\mathcal{P}_s, t(x_S); x) - R(\mathcal{P}_s, t_0(x_S); x)$$
$$\geq \sum_{S \in A_2} \Big[G(|F_x(t_0(x_S)) - \frac{N+1}{2N}| - 1/N) - G(|F_x(t_0(x_S)) - \frac{N+1}{2N}|)$$

(5.13)
$$+ G(|F_x(t_0(x_{S'})) - \frac{N+1}{2N}| + 1/N) - G(|F_x(t_0(x_{S'})) - \frac{N+1}{2N}|) \Big] \mathcal{P}_s(S) \geq 0,$$

where the first inequality holds because we have neglected some summands of the type appearing in the last line of (5.13) that are all in A_1 and are positive since G is increasing. The second inequality follows by convexity of G. $\qquad\square$

The next result compares the maximum risk of linear estimators, which in general are not invariant, including the sample mean which is not covered by Theorem 5.3, and the best invariant estimators Y_{j^*}.

THEOREM 5.4.   *Under the sampling design $\mathcal{P}_s$, the best invariant estimator of the k-th quantile, $t_0 = Y_{j^*}$ with j^* defined in (4.4), is minimax among (symmetric, nonrandomized) estimators that are convex com-*

binations of the type $t_w(\mathbf{Y}) = \sum_{i=1}^{n} w_i Y_i$; that is, $\sup_x R(\mathcal{P}_s, t_w(x_S); x) \geq R(\mathcal{P}_s, t_0(x_S); x) \quad \forall x \in \Upsilon$.

PROOF. By Lemma 5.3 it suffices to show that for some $x \in \Upsilon$ we have

$$(5.14) \qquad R(\mathcal{P}_s, t_w(x_S); x) \geq R(\mathcal{P}_s, t_0(x_S); x),$$

and we show it for x constructed as follows. Let $w = w_k < 1$ (the case $w_k = 1$ is trivial) be the first non zero among $w_1, \ldots, w_n$, and set $x_i = f(i) := 1 - w^i, i = 1, \ldots, N$, and $x = (x_1, \ldots, x_N)$.

We claim that for any $S = \{i_1, \ldots, i_n\}$ we have for the above x, $Y_k = x_{i_k}$, and $x_{i_k} \leq t_w(x_S) < x_{i_k+1}$; this is equivalent to proving that for any $1 \leq i_1 < \ldots < i_n \leq N$ we have $f(i_k) \leq \sum_{j=k}^{n} w_j f(i_j) < f(i_k + 1)$. The left-hand side inequality follows by monotonicity of f, and the right-hand side from $\sum_{j=k}^{n} w_j f(i_j) < 1 - w_k w^{i_k} = 1 - w^{i_k+1} = f(i_k + 1)$.

The relation $x_{i_k} \leq t_w(x_S) < x_{i_k+1}$ implies that for any sample of size n from x, the estimator t_w is equivalent to Y_k, which is an invariant estimator, and by Corollary 4.2 the risk of Y_k is not smaller than that of the best invariant estimator Y_{j^*}, and (5.14) follows. $\quad\square$

References.

Agarwal, O.P. (1955) Some minimax invariant procedures for estimating a cumulative distribution function. *Ann.Math.Statist.*, **26** 450–463.

Arnold, B. C., Balakrishnan, N., and Nagaraja, H. N. (1992) *A First Course in Order Statistics.* Wiley, New York.

Basu, D. (1958) On sampling with and without replacement. *Sankhya*, **20** 287–294.

Basu, D. (1971). An essay on the logical foundations of survey sampling, Part 1 (with discussion), in Foundations of Statistical Inference, eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 203-242.

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York.

Birnbaum, Z. W. 1948 On random variables with comparable peakedness. *Ann. Math. Statist.*, **19** 76-81.

Blackwell, D. and Girshick, M. A. (1954) *Theory of Games and Statistical Decisions.* Wiley, New York.

Brown, L. D. (1988) Admissibility in discrete and continious invariant nonparametric estimation problems and in their multinomial analogs. *The Annals of Statistics*, **16** 1567–1593.

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1977) *Foundations of Inference in Survey Sampling.* Wiley, New York.

Cohen, M. P. and Kuo, L. (1985) Minimax sampling strategies for estimating a finite population distribution function. *Statistics & Decisions*, **3** 205–224.

David, H. A. and Nagaraja, N. H. (2003) *Order Statistics*, 3nd edn. Wiley, New York.

Ferguson, T. (1967) *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press Inc. New York and London.

Godambe, V. P. (1968) Bayesian sufficiency in survey sampling. *Ann. Inst. Statist. Math.*, **20** 363–373.

Godambe, V. P., and Thompson, M.E. (1971) The specification of prior knowledge by classes of prior distributions in survey sampling estimation, in *Foundations of Statistical Inference*, eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 243–254.

Graham, R., Rothschild, B. and Spencer, J. H. (1990) *Ramsey Theory*. John Wiley and Sons, New York.

Kiefer, J. (1957) Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.*, **28** 573–601.

Kirschner, H. P. (1976) On the risk-equivalence of two methods of randomization in statistics . *J. Multivariate Analysis*, **6** 159–166.

Lehmann, E. L., and Casella, G. (1998) *Theory of Point Estimation*, 2nd ed. Springer-Verlag, New York.

Malinovsky, Y. (2009). Nonparametric and parametric estimation of quantiles and ordered parameters of small areas, and related problems. PhD thesis, The Hebrew University of Jerusalem.

Malinovsky, Y. and Rinott, Y. (2009) On stochastic orders of absolute value of order statistics in symmetric distributions. *Statistics and Probability Letters*, **79** 2086-2091.

Nelson, D. and Meeden, Y. (2006) Noninformative nonparametric quantile estimation for simple random samples. *Journal of Statistical Planning and Inference*, **136** 53–67.

Rinott, Y. (2009) Some decision-theoretic aspects in finite population sampling. *in Handbook of Statistics 29B; Sample Surveys: Inference and Analysis.* Eds. D. Pfeffermann and C.R. Rao. Elsevier North Holland 253-558.

Roberts, C. P. (2007) *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* 2nd ed. printing Springer-Verlag, New York.

Shaked, M. and Shanthikumar, J.G. (1994) *Stochastic Orders and Their Applications.* Academic Press, Boston.

Stępień-Baran, A. (2009) Minimax invariant estimator of a continuous distribution function under a general loss function. *Metrika*, in press.

Stenger, H. (1979) A minimax approach to randomization and estimation in survey sampling. *The Annals of Statistics*, **7** 395–399.

Uhlmann, W. (1963) Ranggrössen als Schätzfuntionen (in German). *Metrika*, **7** 23–40.

Wald, A. and and Wolfowitz, J. (1951) Two Methods of Randomization in Statistics and the Theory of Games *The Annals of Mathematics*, **53** 581-586

Wilks, S. S. (1962) *Mathematical Statistics*. Wiley, New York.

Yu, Q. and Chow, M. (1991) Minimaxity of the empirical distribution function in invariant estimation. *The Annals of Statistics*, **19** 935–951.

Yu, Q. and Phadia, E. (1992) Minimaxity of the best invariant estimator of a distribution function under the Kolmogorov-Smirnov loss. *The Annals of Statistics*, **20** 2192–2195.

Zieliński, R. (1999) Best equivariant nonparametric estimator of a quantile. *Statistics and Probability Letters*, **45** 79–84.

Division of Epidemiology, Statistics and Prevention Research
National Institute of Child Health and Human Development (NICHD)
6100 Executive Boulevard, Bethesda, MD 20852
U.S.A.

E-mail: yaakov.malinovsky@nih.gov

Department of Statistics
The Hebrew University of Jerusalem
Mount Scopus, Jerusalem 91905
Israel

E-mail: rinott@mscc.huji.ac.il