Minimax risks for distributed estimation of the background in a field of noise sources

M. Madiman, *Member, IEEE,* A. R. Barron, *Senior Member, IEEE,* A. M. Kagan and T. Yu

(Invited Paper for WITS 2008)

***Abstract*—Consider a scenario where some background quantity is to be measured, but the only access to its measurement is through a collection of sensors that observe finite samples of this quantity corrupted by the field of noisy sources in which the sensors are embedded. A model for such a scenario is presented, and the fundamentally best achievable statistical performance for the sensors is studied in terms of minimax risks. Applications are given to design and resource allocation problems in sensor networks whose goal is the distributed estimation of a background.**

***Index Terms*—Distributed background estimation, location parameter, minimax risk.**

I. Introduction

A VERY important problem in astronomy with significant implications for cosmology is the measurement of the cosmic microwave background radiation. This is, in physical parlance, radiation left over from the very hot gases that pervaded the universe soon after the Big Bang, and its precise measurement can help to test various key hypotheses about cosmological constants and origins. A feature of the measurements that are made, however, is that they include not just measurement noise, but also the effects of other sources of similar radiation in the universe, such as intergalactic clouds, whose distribution in different directions is uneven. This kind of situation, where there is a background parameter to be estimated in the presence of a bunch of sources that may be considered noise for our purposes, arises in other applications as well. We present a simplified model for such a scenario, discuss the measurement of the background parameters using a collection of sensors, and then perform a theoretical analysis of the best distributed estimation performance achievable for this model.

The central results of our work are inequalities relating how well different sensors, that have access to different portions of the data, can do in terms of estimating an underlying, common location parameter. Let us motivate this problem of distributed estimation (of, for instance, a mean) with an example different from the one already discussed. Suppose one is interested in measuring temperature, chemical concentration, or other random variables that are geographically distributed. In some scenarios, it is the geographical variation of the distributions that are of interest; in others, it is common parameters underlying the entire distribution. An example of the latter is when one wants to detect deviations of the common parameter from an allowed parameter range. Typically these kinds of deviations happen slowly by drift of the underlying parameter over time. It is therefore fair to assume for purposes of analysis that the underlying parameter remains constant over short periods of time (and therefore over a certain sample size for observations); however, it is rather unrealistic to assume that these sample sizes can be taken to be infinitely large. Thus one is interested in the accuracy of estimates that can be made of the parameter by sensors using *finite* (and possibly small) samples of measurements.

Our model for sensor networks thus consists of the following components:

- A field of N "sources", that produce data streams of size T. We assume there is spatial independence, that is, the data streams produced by different sources are independent of each other. However, arbitrary temporal dependence is allowed, so that the sample produced by a single source can come from any joint distribution.
- A network of sensors, each corresponding to a subset $\mathbf{s} \in \mathcal{C}$, where $\mathcal{C}$ is a collection of subsets of $[N] = \{1, 2, \ldots, N\}$. Thus each choice of a collection $\mathcal{C}$ corresponds to a particular *sensor network configuration*.
- The $\mathbf{s}$-sensor has access to a combination of the data coming from the sources in $\mathbf{s}$. For the purposes of this paper, we assume that what is measured is the background parameter corrupted additively by the simplest non-trivial combination of the data from the sources– namely, the sum. It is also convenient to assume that there is no additional noise from the process of measurement. (Note that if there is such additional noise, and if its probability distribution is known, we can treat it just as another source in the field of sources).

Our goal is not to come up with heuristically motivated algorithms for optimal distributed estimation. Instead it is to first understand the fundamental limits of distributed estimation, before we augment the model with communication and computation constraints that are important considerations for real-life sensor networks. Indeed, we present the first rigorous analysis of the best possible performance of arbitrary sensor network configurations based on finite sets of observations by using a decision-theoretic framework based on minimax risks. Although our model is a toy model because we ignore communication and computation constraints, it is a first step to the rigorous analysis of fundamentally optimal distributed estimation in more complex settings.

This note is organized as follows. For any sensor (cor-

MM and ARB are with the Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA. Email: mokshay.madiman@yale.edu, andrew.barron@yale.edu

AMK and TY are with the Department of Mathematics, University of Maryland, College Park, MD 20742, USA. Email: amk@math.umd.edu, yuth@math.umd.edu

responding to a set of sources) and any sample size, there is an associated number– the minimax risk– which captures the smallest mean squared error of estimation that can be achieved uniformly over all possible parameter values. Section II describes our main results, which focus on the minimax risks associated with distributed estimation of a background parameter. Section III contains a proof of the main technical inequality, which provides a comparison between the minimax risks achieved by the sensors in an arbitrary sensor network configuration, and the minimax risk achieved by a single sensor that is exposed to all noise sources combined. Section IV uses this inequality to prove some results on sensor network design and resource allocation.

II. Main Results

First we describe a model for distributed estimation of a background based on the considerations discussed in Section I. Recall that $[N]$ denotes the index set $\{1, 2, \ldots, N\}$.

A precise description of the components of the model is as follows:

1) There are N sources, indexed by the set $[N]$, and these sources are independent of each other (i.e., the data streams produced by the sources are independent of each other).
2) For each $i \in [N]$, source i produces the data stream $\mathbf{X}_i = (X_{i,1}, \ldots, X_{i,T})$ from some known joint distribution F_i on $\mathbb{R}^T$. The distribution F_i is arbitrary, except for the reasonable assumption that its covariance matrix is finite. It is convenient to think of the data stream $\mathbf{X}_i$ as data produced over time, and T as the number of time periods for which data is available.
3) The background quantity θ is some unknown real number. From a statistical point of view, we treat it as a *parameter*.
4) The $\mathbf{s}$-sensor observes the subset sums $\mathbf{Y_S} = (Y_{\mathbf{S},1}, Y_{\mathbf{S},2}, \ldots, Y_{\mathbf{S},T})$, where
$$Y_{\mathbf{S},t} = \theta + \sum_{i \in \mathbf{S}} X_{i,t}.$$
Note that for any subset $\mathbf{s} \subset [N]$, the subset sum distribution $F_{\mathbf{S}}$ is obtained by the convolution of the distributions indexed by the elements of $\mathbf{s}$, thus we write $F_{\mathbf{S}} = *_{i \in \mathbf{S}} F_i$. Then, in statistical language, $\mathbf{Y_S}$ is a sample from the location family $F_{\mathbf{S}}(x_1 - \theta, \ldots, x_T - \theta)$ generated by the subset sum distribution $F_{\mathbf{S}}$.
5) From the T observations in $\mathbf{Y_S}$, the $\mathbf{s}$-sensor constructs an estimate $\tilde{\theta}_{\mathbf{S}}(\mathbf{Y_S})$ of the parameter θ.

Our goal is to relate the statistical properties of different estimates of a parameter obtained by users who have access to different sets of observations, in accordance with this model. The goodness of an estimator is measured by comparing to the "best possible estimator in the worst case", i.e., by comparing the risk (or mean square error) of the given estimator with the minimax risk. The minimax risk achievable by the $\mathbf{s}$-user is

$$R_T(\mathbf{s}) = \min_{\text{all estimators } \tilde{\theta}_{\mathbf{S}}} \max_{\theta} \mathbb{E}[(\tilde{\theta}_{\mathbf{S}}(\mathbf{Y_S}) - \theta)^2].$$

For location problems, assuming that there exists an estimator with finite risk, this risk is achievable, since Girshick and Savage [1] proved that the Pitman estimator is minimax. Here by Pitman estimator we mean the estimator with minimum mean square error among all equivariant estimators of location (see, eg., [3], for definitions and details).

A main result of this note is the following inequality relating the minimax risks achievable by the $\mathbf{s}$-users from the class $\mathcal{C}$ to the minimax risk achievable by the $[N]$-user, i.e, one who only sees observations from the location family generated by the total sum.

Theorem 1: Suppose $F_1, \ldots, F_N$ have finite covariance matrices. Let $\mathcal{C}$ be a regular collection in the sense that each index $i \in [N]$ appears in exactly r sets in $\mathcal{C}$. Then, for any sample size $T \geq 1$,

$$R_T([N]) \geq \frac{1}{r} \sum_{\mathbf{S} \in \mathcal{C}} R_T(\mathbf{s}).$$

Our approach to minimax risks for location families is based on the fact that the Pitman estimator mentioned above is minimax in this setting. In the case where $\mathcal{C}$ is simply the collection of singleton sets $\{1\}, \ldots, \{N\}$, and each source is producing a data stream of T i.i.d. observations, Theorem 1 states that

$$\mathrm{var}(\hat{\theta}^{(T)}_{[N]}) \geq \sum_{i \in [N]} \mathrm{var}(\hat{\theta}^{(T)}_i),$$

where $\hat{\theta}^{(T)}_i$ is the Pitman estimator based on i.i.d. observations from source i, and $\hat{\theta}^{(T)}_{[N]}$ is the Pitman estimator based on i.i.d. observations from the location family generated by $F_{[N]}$. This special case was proved by Kagan [2], although not interpreted there in this way.

We now discuss a result comparing several different sensor network configurations. Note that for the sensor network configuration corresponding to a collection $\mathcal{C}$ of subsets of $[N]$, a reasonable figure of merit is the average minimax risk per element observed, i.e., the quantity

$$\frac{1}{|\mathcal{C}|} \sum_{\mathbf{S} \in \mathcal{C}} \frac{R_T(\mathbf{s})}{|\mathbf{s}|}.$$

This is more appropriate than simply the average minimax risk (without the normalization by $|\mathbf{s}|$ inside the summation), which does not take into account the advantage of sensitive measurements corresponding to smaller set sizes.

Theorem 2: [Hierarchy for symmetric collections] For the collection $\mathcal{C}_k$ of all subsets of size k, let

$$A_k = \frac{1}{\binom{N}{k}} \sum_{\mathbf{S} \in \mathcal{C}_k} \frac{R_T(\mathbf{s})}{k}$$

be the average minimax risk per element observed. Then

$$A_1 \leq A_2 \leq \ldots \leq A_{N-1} \leq A_N.$$

What Theorem 2 says is that even taking into account the advantage of sensitive sensors, using N sensitive sensors that

pick up the individual sources is still better than using $\binom{N}{2}$ rough sensors picking up all pairwise sums. Although the usefulness of this statement is limited because we have ignored communication and data fusion aspects, it still gives some insight into the design question for sensor networks, and in particular tells us something about the tradeoff between the sensitivity and the number of sensors. As far as we are aware, this is the first such rigorous result in sensor network theory in the robust framework of minimax risks.

Theorem 2 is particularly striking in the case where all the sources have the same distribution, i.e., $F_1 = F_2 = \ldots = F_N$. In this case, it implies for instance that

$$\frac{R_T([N])}{N} \geq \frac{R_T([N-1])}{N-1}.$$

Thus we have a direct relationship between the efficacy of different individual sensors; indeed, under the assumption that all sources are probabilistically identical, sensors exposed to fewer sources are always better (in the sense of minimax risk per element observed) than sensors exposed to more of them.

We now move on to a question of resource allocation. Suppose we can give *variance permissions* V_i for each source, i.e., the $\mathbf{s}$-sensor is only allowed an estimator with variance less than or equal to $\sum_{i\in\mathbf{s}} V_i$. Clearly for the $[N]$-sensor to be feasible under the given variance permissions, it needs to be able to estimate the location parameter with mean square error uniformly not more than $\sum_{i\in[N]} V_i$. So the smallest total variance we can hope to allot is this sum of all the V_i's. However it is not at all obvious that one can find such a variance allotment (with no total wasted variance) for which *any other sensor network configuration is also feasible*. This question is solved in the next result.

Theorem 3: [Resource Allocation] Let V_i be the variance permission associated with source i. In other words, the $\mathbf{s}$-sensor is allowed a worst-case mean squared error of estimation of at most $\sum_{i\in\mathbf{s}} V_i$. For an arbitrary sensor configuration to be feasible, the definition of minimax risks means that one needs

$$\sum_{i\in\mathbf{S}} V_i \geq R_T(\mathbf{s}) \qquad \text{for each } \mathbf{s} \subset [N].$$

Under this constraint, it is possible to allot variance permissions to all sources in such a way that there is no wasted total variance, i.e., $\sum_{i\in[N]} V_i = R_T([N])$.

These results are proved in the following sections.

III. Relations between minimax risks

We need to define the notion of a *fractional partition* for a collection $\mathcal{C}$ of subsets of $[N]$. A set $\beta = \{\beta(\mathbf{s}) : \mathbf{s} \in \mathcal{C}\}$ of non-negative real numbers is called a fractional partition for $\mathcal{C}$ if

$$\sum_{\mathbf{S}\ni i, \mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) = 1 \tag{1}$$

for each i in $[N]$.

If the numbers $\beta(\mathbf{s})$ are constrained to only take the values 0 and 1, then the condition above entails that exactly one set in $\mathcal{C}$ contains i, so that the sets $\mathbf{s} \in \mathcal{C}$ are pairwise disjoint and fill out the set $[N]$, forming a partition of $[N]$. We may interpret a fractional partition as a "partition" of $[N]$ using sets in $\mathcal{C}$, each of which contains only a fractional piece (namely, $\beta_{\mathbf{S}}$) of the elements in that set. Another way of saying this is to note that for any real numbers a_i,

$$\begin{aligned}\sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) \sum_{i\in\mathbf{S}} a_i &= \sum_{i\in[N]} a_i \sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) \mathbf{1}_{\{i\in\mathbf{S}\}} \\ &= \sum_{i\in[N]} a_i.\end{aligned} \tag{2}$$

We now present a generalization of Theorem 1, and sketch its proof.

Theorem 4: Suppose all the sources have finite covariance matrices. Then for any sample size $T \geq 1$, and for any fractional partition β,

$$R_T([N]) \geq \sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) R_T(\mathbf{s}).$$

Proof: Since the sample size T is fixed for the rest of the paper, we simply use $\hat{\theta}$ rather than $\hat{\theta}^{(T)}$ to denote Pitman estimators. For any $\mathbf{s}$, the Pitman estimator for θ based on the observations $\mathbf{Y_S}$ can be written as

$$\hat{\theta}_{\mathbf{S}}(\mathbf{Y_S}) = \bar{Y}_{\mathbf{S}} - \mathbb{E}(\bar{Y}_{\mathbf{S}} | \mathcal{G}_{\mathbf{S}}), \tag{3}$$

where

$$\bar{Y}_{\mathbf{S}} = \frac{1}{T} \sum_{i=1}^{T} Y_{\mathbf{S},t} \tag{4}$$

is the sample mean,

$$\mathcal{G}_{\mathbf{S}} = \sigma(Y_{\mathbf{S},1} - \bar{Y}_{\mathbf{S}}, \ldots, Y_{\mathbf{S},T} - \bar{Y}_{\mathbf{S}}) \tag{5}$$

is the σ-algebra generated by the residuals, and $\mathbb{E}$ stands for the expectation taken at $\theta = 0$. This expression for the Pitman estimator is based on the theory of equivariant estimation, as described, eg., in [3].

By the finite variance assumption, the sample mean is a square-integrable function, so that one may interpret (3) in terms of an orthogonal projection. Associated with the projection is a Pythagorean identity, namely,

$$\mathrm{var}(\hat{\theta}_{\mathbf{S}}) = \mathrm{var}(\bar{Y}_{\mathbf{S}}) - \mathrm{var}\big(\mathbb{E}(\bar{Y}_{\mathbf{S}} | \mathcal{G}_{\mathbf{S}})\big), \tag{6}$$

which follows from the uncorrelatedness of the projection $E(\bar{Y}_{\mathbf{S}} | \mathcal{G}_{\mathbf{S}})$ and $\hat{\theta}_{\mathbf{S}}$. First we note that

$$\begin{aligned}\sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) \mathrm{var}(\bar{Y}_{\mathbf{S}}) &= \sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) \sum_{j\in\mathbf{S}} \mathrm{var}(\bar{Y}_{\{j\}}) \\ &= \sum_{j\in[N]} \mathrm{var}(\bar{Y}_{\{j\}}) \\ &= \mathrm{var}(\bar{Y}_{[N]}),\end{aligned} \tag{7}$$

where we used the identity (2) for a fractional partition. We will now show that

$$\mathbb{E}\Big[\mathbb{E}[\bar{Y}_{[N]} | \mathcal{G}_{[N]}]\Big]^2 \leq \sum_{\mathbf{S}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}\Big[\mathbb{E}[\bar{Y}_{\mathbf{S}} | \mathcal{G}_{\mathbf{S}}]\Big]^2. \tag{8}$$

Combining (7) and (8), and keeping in mind that variances and second moments coincide because the expectation is taken with respect to $\theta = 0$, one obtains

$$\mathrm{var}(\bar{Y}_{[N]}) - \mathrm{var}\big(\mathbb{E}(\bar{Y}_{[N]}|\mathcal{G}_{[N]})\big) \geq \sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \Big[\mathrm{var}(\bar{Y}_{\mathbf{S}}) - \mathrm{var}\big(\mathbb{E}(\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}})\big) \Big]. \tag{9}$$

Then the variance decomposition (6) gives

$$\mathrm{var}(t_{[N]}) \geq \sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathrm{var}(t_{\mathbf{S}}), \tag{10}$$

which proves the inequality of Theorem 4 due to the minimaxity of the Pitman estimator.

To demonstrate (8), we observe that

$$\begin{aligned} \sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}\Big[\mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}]\Big]^2 &\overset{(a)}{\geq} \mathbb{E}\Big[\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}]\Big]^2 \\ &= \mathbb{E}\mathbb{E}\Big[\Big\{\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}]\Big\}^2 \Big| \mathcal{G}_{[N]}\Big] \\ &\overset{(b)}{\geq} \mathbb{E}\Big[\mathbb{E}\Big\{\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}] \Big| \mathcal{G}_{[N]}\Big\}\Big]^2, \end{aligned} \tag{11}$$

where (a) follows by the variance drop lemma of Madiman and Barron [4], and (b) from the Cauchy-Schwarz inequality. Also, the independence of the random variables implies that $\mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}] = \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}, \mathcal{G}_{\mathbf{S}^c}]$, so that

$$\begin{aligned} &\mathbb{E}\Big\{\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}] \Big| \mathcal{G}_{[N]}\Big\} \\ &= \sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}\Big\{\mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}, \mathcal{G}_{\mathbf{S}^c}] \Big| \mathcal{G}_{[N]}\Big\} \\ &\overset{(c)}{=} \sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{[N]}] \\ &= \mathbb{E}\Big\{\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \bar{Y}_{\mathbf{S}} \Big| \mathcal{G}_{[N]}\Big\} \\ &\overset{(d)}{=} \mathbb{E}\big\{\bar{Y}_{[N]} \Big| \mathcal{G}_{[N]}\big\}, \end{aligned} \tag{12}$$

where (c) follows from the fact that the coarser σ-algebra wins, and (d) follows by applying the identity (2) to the numbers $\bar{Y}_{\{j\}}$. Thus

$$\mathbb{E}\Big[\mathbb{E}\Big\{\sum_{\mathbf{s}\in\mathcal{C}} \beta(\mathbf{s}) \mathbb{E}[\bar{Y}_{\mathbf{S}}|\mathcal{G}_{\mathbf{S}}] \Big| \mathcal{G}_{[N]}\Big\}\Big]^2 = \mathbb{E}\Big[\mathbb{E}\big\{\bar{Y}_{[N]}\big|\mathcal{G}_{[N]}\big\}\Big]^2,$$

which completes the proof. ∎

Remark 1: It is easy to see that Theorem 1 follows from Theorem 4. Indeed, if $r(i) = |\{\mathbf{s} \in \mathcal{C} : i \in \mathbf{s}\}|$ is the *degree* of i, the regularity condition on $\mathcal{C}$ means that $r(i) = r$ for each $i \in [N]$, so that $\sum_{\mathbf{S}\ni i, \mathbf{S}\in\mathcal{C}} 1 = r$ and the coefficients $\beta(\mathbf{s}) = \frac{1}{r}$ form a fractional partition.

Remark 2: Although it is not relevant to the focus of this paper, it is worth mentioning that Theorems 1 and 4 are generalizations of the subset sum Fisher information inequalities of [4].

IV. Design and Resource Allocation Issues

Consider the special case of Theorem 1 corresponding to the collection of leave-one-out subsets, i.e., to the collection

$$\mathcal{C}_{N-1} = \{\mathbf{s} : |\mathbf{s}| = N - 1\}.$$

Then Theorem 1 reads

$$R_T([N]) \geq \frac{1}{N-1} \sum_{\mathbf{s}\in\mathcal{C}_{N-1}} R_T(\mathbf{s})$$

so that

$$\frac{R_T([N])}{N} \geq \frac{1}{N} \sum_{\mathbf{s}\in\mathcal{C}_{N-1}} \frac{R_T(\mathbf{s})}{N-1}. \tag{13}$$

This proves the inequality $A_N \geq A_{N-1}$, which is part of Theorem 2. The rest is proved recursively by repeated application of inequality (13) to each of the summands appearing in the right side of (13).

To prove Theorem 3, consider the linear program

$$\begin{aligned} &\text{Maximize } \textstyle\sum_{\mathbf{s}\subset[N]} \beta(\mathbf{s}) R_T(\mathbf{s}) \\ &\text{subject to} \quad \beta(\mathbf{s}) \geq 0 \text{ for each } \mathbf{s} \subset [N] \\ &\text{and} \quad \textstyle\sum_{\mathbf{s}\subset[N], \mathbf{s}\ni j} \beta(\mathbf{s}) = 1 \text{ for each } j \in [N]. \end{aligned}$$

The dual problem is easily obtained:

$$\begin{aligned} &\text{Minimize } \textstyle\sum_{j\in[N]} V_j \\ &\text{subject to} \quad \textstyle\sum_{j\in\mathbf{s}} V_j \geq R_T(\mathbf{s}) \text{ for each } \mathbf{s} \subset [N]. \end{aligned}$$

If p^* and d^* denote the primal and dual optimal values, duality theory tells us that $p^* = d^*$. But Theorem 4 implies that $p^* \leq R_T([N])$, since by setting $\beta(\mathbf{s}) = 0$ for some subsets $\mathbf{s} \subset [N]$, fractional partitions using arbitrary collections of sets can be thought of as fractional partitions using the full power set $2^{[N]}$. Hence it must be true that $d^* \leq R_T([N])$. Thus there exists a point $(V_1, \ldots, V_N)$ in the feasible region of the dual problem such that the sum $V_1 + \ldots + V_N = R_T([N])$.

References

[1] M. A. Girshick and L. J. Savage. Bayes and minimax estimates for quadratic loss functions. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 53–73, Berkeley and Los Angeles, 1951. University of California Press.

[2] A. Kagan. An inequality for the Pitman estimators related to the Stam inequality. *Sankhyā Ser. A*, 64:281–292, 2002.

[3] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.

[4] M. Madiman and A.R. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory*, 53(7):2317–2329, July 2007.