A NORMAL HIERARCHICAL MODEL AND MINIMUM CONTRAST ESTIMATION FOR RANDOM INTERVALS

By Yan Sun* and Dan Ralescu†

Utah State University and University of Cincinnati

Many statistical data are imprecise due to factors such as measurement errors, computation errors, and lack of information. In such cases, data are better represented by intervals rather than by single numbers. Existing methods for analyzing interval-valued data include regressions in the metric space of intervals and symbolic data analysis, the latter being proposed in a more general setting. However, there has been a lack of literature on the distribution-based inferences for interval-valued data. In an attempt to fill this gap, we extend the concept of normality for random sets by Lyashenko (1983) and propose a normal hierarchical model for random intervals. In addition, we develop a minimum contrast estimator (MCE) for the model parameters, which we show is both consistent and asymptotically normal. Simulation studies support our theoretical findings, and show very promising results. Finally, we successfully apply our model and MCE to a real dataset.

1. Introduction. In classical statistics, it is often assumed that the outcome of an experiment is precise and the uncertainty of observations is solely due to randomness. Under this assumption, numerical data are represented as collections of real numbers. In recent years, however, there has been increased interest in situations when exact outcomes of the experiment are very difficult or impossible to obtain, or to measure. The imprecise nature of the data thus collected is caused by various factors such as measurement errors, computational errors, loss or lack of information. Under such circumstances and, in general, any other circumstances such as grouping and censoring, when observations cannot be pinned down to single numbers, data are better represented by intervals. Practical examples include interval-valued stock prices, oil prices, temperature data, medical records, mechanical measurements, among many others.

The effort in the literature to analyze interval-valued data, while still at its early stage, shows promising advances. The earliest attempt probably dates

*Corresponding author.

†Supported by a Taft Research Grant.

AMS 2000 subject classifications: Primary 62H10, 62H12; secondary 52A22

Keywords and phrases: random intervals, Normality, hierarchical, Choquet functional, minimum contrast estimator, strong consistency, asymptotic normality.

back to the 1990s, when Diamond published his paper on the least squares fitting of compact set-valued data and considered interval-valued input and output as a special case (see Diamond, 1990). Due to the embedding theorems started by Brunn and Minkowski and later refined by Rådström (see Rådström 1952) and Hörmander (see Hörmander 1954), $\mathcal{K}(\mathbb{R}^n)$, the space of all nonempty compact convex subsets of $\mathbb{R}^n$, is embedded into the Banach space of support functions. Diamond (1990) defined a L_2 metric in this Banach space of support functions, and found the regression coefficients by minimizing the L_2 metric of the sum of residuals. This idea was further studied in Gil et al. (2002), where the L_2 metric was replaced by a generalized metric on the space of nonempty compact intervals, called "W-distance", proposed earlier by Körner and Nather (1998). Separately, Billard and Diday (2003) introduced the central tendency and dispersion measures and developed the symbolic interval data analysis based on those. (See also Carvalho et al., 2004.) However, none of the existing literature considered distributions of the random intervals and the corresponding statistical methods.

It is well known that normality plays an important role in classical statistics. But the normal distribution for random sets remained undefined for a long time, until the 1980s when the concept of normality was first introduced for compact convex random sets in the Euclidean space by Lyashenko (1983). It is especially useful in deriving limit theorems for random sets. See, Puri et al. (1986), Norberg (1984), among others. Since a compact convex set in $\mathbb{R}$ is a closed bounded interval, by the definition of Lyashenko (1983), a normal random interval is simply a displacement of a fixed closed bounded interval. From the point of view of statistics, this is not enough to fully capture the randomness of a general random interval.

In this paper, we extend the definition of normality given by Lyashenko (1983) and propose a normal hierarchical model for random intervals. With one more degree of freedom on "shape", our model conveniently captures the entire randomness of random intervals via a few parameters. Therefore, it adds to the literature the possibility of distribution-based inference methods for interval-valued data. Especially, conditioning on the first hierarchy, our normal hierarchical random interval is exactly the normal random interval defined by Lyashenko (1983). This could be a very useful property in view of the limit theorems. In addition, with certain choices of the distributions, a linear combination of our normal hierarchical random intervals follows the same normal hierarchical distribution. An immediate consequence of this property is the possibility of a factor model for multi-dimensional random intervals, based on our normal hierarchical distribution, as the "factor" will have the same distribution as the original intervals.

To estimate the parameters and make inferences for our Normal hierarchical model, we propose a minimum contrast estimator (MCE) based on the hitting function of the random interval. We show that under certain conditions the MCE satisfies a strong consistency and asymptotic normality. A simulation study is carried out for one specific distribution, and the results are consistent with our theorems. We apply our model to analyze a daily temperature range data and, in this context, we have derived interesting and promising results.

The rest of the paper is organized as follows. Section 2 formally defines our Normal hierarchical model and discusses its statistical properties. Section 3 introduces a minimum contrast estimator for the model parameters, and presents its asymptotic properties. A simulation study is reported in Section 4, and a real data application is demonstrated in Section 5. We give concluding remarks in Section 6. Proofs of the theorems are presented in Section 7. Useful lemmas and other proofs are deferred to the Appendix.

2. The Normal hierarchical model.

2.1. *Definition.* Let $(\Omega, \mathcal{L}, P)$ be a probability space. Denote by $\mathcal{K}$ the collection of all non-empty compact subsets of $\mathbb{R}^d$. A random compact set is a Borel measurable function $A : \Omega \to \mathcal{K}$, $\mathcal{K}$ being equipped with the Borel σ-algebra induced by the Hausdorff metric. If $A(\omega)$ is convex for almost all ω, then A is called a random compact convex set. (See Molchanov 2005, p.21, p.102.) Denote by $\mathcal{K}_{\mathcal{C}}$ the collection of all compact convex subsets of $\mathbb{R}^d$. By Theorem 1 of Lyashenko (1983), a compact convex random set A in the Euclidean space $\mathbb{R}^d$ is Gaussian if and only if A can be represented as the Minkowski addition of a fixed compact convex set M and a d-dimensional normal random vector ϵ, i.e.

$$A = M + \{\epsilon\} . \tag{1}$$

As pointed out in Lyashenko (1983), the Gaussian random set defined above is especially useful in view of the limit theorems discussed earlier in Lyashenko (1979). That is, if the conditions in those theorems are satisfied and the limit exists, then it is Gaussian in the sense of (1). Puri et al. (1986) extended these results to a separable Banach space.

In the following, we will restrict ourselves to compact convex random sets in $\mathbb{R}^1$, that is, bounded closed random intervals. They will be called random intervals for ease of presentation.

According to (1), a random interval A is Gaussian if and only if A is representable in the form

$$A = I + \{\epsilon\} , \tag{2}$$

where I is a fixed bounded closed interval and ϵ is a normal random variable. Obviously, such a random interval is simply a displacement of a fixed interval, so it is not enough to fully capture the randomness of a general random interval. In order to model the randomness of both the location and the "shape" (length), we propose the following Normal hierarchical model for random intervals:

$$A = I + \{\epsilon\}, \tag{3}$$
$$I = \eta I_0, \tag{4}$$

where η is another random variable and I_0 is a fixed interval in $\mathbb{R}$. Here, the product ηI_0 is in the sense of scalar multiplication of a real number and a set. Let $\lambda(\cdot)$ denote the Lebesgue measure of a compact convex subset of $\mathbb{R}^d$, which in the case $d = 1$ is the length of an interval. Then,

$$\lambda(A) = \lambda(\epsilon + \eta I_0) = \lambda(\eta I_0) = |\eta|\, \lambda(I_0).$$

That is, η is the parameter that models the length of A. In particular, if $\eta \to 0$, then A reduces to a normal random variable.

An interesting property of the Normal hierarchical random interval is that its linear combination is still a Normal hierarchical random interval. This is seen by simply observing that

$$\sum_{i=1}^{n} a_i A_i = \sum_{i=1}^{n} a_i \left(\epsilon_i + \eta_i I_0\right) = \sum_{i=1}^{n} a_i \epsilon_i + I_0 \left(\sum_{i=1}^{n} a_i \eta_i\right), \tag{5}$$

for arbitrary constants $a_i, i = 1, \cdots, n$, where "+" denotes the Minkowski addition. This is very useful in developing a factor model for the analysis of multiple random intervals. Especially, if we assume $\eta_i \sim N(\mu_i, \sigma_i^2), i = 1, \cdots, n$, then the "factor" $\sum_{i=1}^{n} a_i A_i$ has exactly the same distribution as the original random intervals. We will elaborate more on this issue in section 4.

Without loss of generality, we can assume in the model (3)-(4) that $E\epsilon = 0$. We will make this assumption throughout the rest of the paper.

2.2. *Model properties.* The fundamental theory of random sets was developed in the 1960s and 1970s. See, e.g., Matheron (1967), Matheron (1975), and Kendall (1974). According to the Choquet theorem (Molchanov 2005, p.10), the distribution of a random closed set (and random compact convex set as a special case) A, is completely characterized by the hitting function T defined as:

$$T(K) = P(K \cap A \neq \emptyset), \quad \forall K \in \mathcal{K}_{\mathcal{C}}. \tag{6}$$

T is also called the Choquet capacity functional. For compact convex sets, there is another functional, containment functional $C(K) = P(A \subset K)$, $\forall K \in \mathcal{K}_{\mathcal{C}}$, which also uniquely determines the distribution. But we are not considering $C(K)$ in this paper.

Writing $I_0 = [a_0, b_0]$ with $a_0 \leq b_0$, the Normal hierarchical random interval in (3)-(4) has the following hitting function: $\forall [a, b] \in \mathcal{K}_{\mathcal{C}}$,

$$\begin{aligned} & T_A([a,b]) \\ = \;& P([a,b] \cap A \neq \emptyset) \\ = \;& P([a,b] \cap A \neq \emptyset, \eta \geq 0) + P([a,b] \cap A \neq \emptyset, \eta < 0) \\ = \;& P(a - \eta b_0 \leq \epsilon \leq b - \eta a_0, \eta \geq 0) + P(a - \eta a_0 \leq \epsilon \leq b - \eta b_0, \eta < 0). \end{aligned}$$

The expected value of a compact convex random set A is defined by the Aumann integral of a set-valued function (see Aumann 1965, Artstein and Vitale 1975) as

$$EA = \{E\xi : \xi \in A \text{ almost surely}\}.$$

In particular, the Aumann expectation of a random interval A is given by

$$EA = [EA_l, EA_u], \tag{7}$$

where A_l and A_u are the lower and upper bound of A respectively. Therefore, the Aumann expectation of the Normal hierarchical random interval A is

$$\begin{aligned} EA &= E(\epsilon + \eta I_0) = E\epsilon + E(\eta I_0) = E(\eta I_0) \\ &= E\left\{[a_0\eta, b_0\eta] I_{(\eta \geq 0)} + [b_0\eta, a_0\eta] I_{(\eta < 0)}\right\} \\ &= E\left[a_0\eta I_{(\eta \geq 0)} + b_0\eta I_{(\eta < 0)}, b_0\eta I_{(\eta \geq 0)} + a_0\eta I_{(\eta < 0)}\right] \\ &= \left[a_0 E\eta_+ + b_0 E\eta_-, b_0 E\eta_+ + a_0 E\eta_-\right], \end{aligned}$$

where

$$\begin{aligned} \eta_+ &= \eta I_{(\eta \geq 0)}, \\ \eta_- &= \eta I_{(\eta < 0)}. \end{aligned}$$

Notice that η_+ can be interpreted as the positive part of η, but η_- is not the negative part of η, as $\eta_- < 0$ when $\eta < 0$.

The variance of a compact convex random set A in $\mathbb{R}^d$ is defined via its support function. (See Körner 1995, 1997.) Let S^{d-1} be the unit sphere in $\mathbb{R}^d$, and let μ be the normalized $(d-1)$-dimensional Lebesgue measure on S^{d-1}, i.e., $\mu(S^{d-1}) = 1$. The support function $s_A(\cdot)$ of A is defined as

$$s_A(u) = \sup_{a \in A} < u, a >, \forall u \in S^{d-1}.$$

A compact convex set A corresponds uniquely to its support function $s_A(\cdot)$. See Schneider (1993, p.37) for example. Let $\|\cdot\|_2$ denote the L_2-metric in the space of Lebesgue square integrable functions on S^{d-1}. The δ_2 metric in $\mathcal{K}_\mathcal{C}$ is defined by

$$\delta_2(A,B) = \|s_A - s_B\|_2 = \left\{ d \int_{S^{d-1}} |s_A(u) - s_B(u)|^2 \mu(du) \right\}^{\frac{1}{2}}, \forall A, B \in \mathcal{K}_\mathcal{C}.$$

Then the variance of A is defined as

$$Var(A) = E\delta_2^2(A, EA), \tag{8}$$

where EA is the Aumann expectation defined in (7). In the special case when $d = 1$, it is shown by straightforward calculations that

$$Var(A) = \frac{1}{2} Var(A_l) + \frac{1}{2} Var(A_u), \tag{9}$$

for a random interval A. See, for example, Körner (1995). For the Normal hierarchical random interval A,

$$\begin{aligned}
& Var(A_l) \\
= \; & Var\left(\epsilon + a_0\eta_+ + b_0\eta_-\right) \\
= \; & E\left(\epsilon + a_0\eta_+ + b_0\eta_-\right)^2 - \left[E\left(\epsilon + a_0\eta_+ + b_0\eta_-\right)\right]^2 \\
= \; & E\epsilon^2 + a_0^2 Var(\eta_+) + b_0^2 Var(\eta_-) \\
& +2\left(a_0 E\epsilon\eta_+ + b_0 E\epsilon\eta_- - a_0 b_0 E\eta_+ E\eta_-\right),
\end{aligned}$$

and,

$$\begin{aligned}
& Var(A_u) \\
= \; & Var\left(\epsilon + b_0\eta_+ + a_0\eta_-\right) \\
= \; & E\left(\epsilon + b_0\eta_+ + a_0\eta_-\right)^2 - \left[E\left(\epsilon + b_0\eta_+ + a_0\eta_-\right)\right]^2 \\
= \; & E\epsilon^2 + b_0^2 Var(\eta_+) + a_0^2 Var(\eta_-) \\
& +2\left(b_0 E\epsilon\eta_+ + a_0 E\epsilon\eta_- - a_0 b_0 E\eta_+ E\eta_-\right).
\end{aligned}$$

The variance of A is then found to be

$$\begin{aligned}
Var(A) &= \frac{1}{2} Var(A_l) + \frac{1}{2} Var(A_u) \\
&= E\epsilon^2 + \frac{1}{2}\left(a_0^2 + b_0^2\right)\left[Var(\eta_+) + Var(\eta_-)\right] \\
&\quad +(a_0 + b_0)E\epsilon\eta - 2a_0 b_0 E\eta_+\eta_-.
\end{aligned}$$

Remark. From the discussion earlier in this section, we see that for the Normal hierarchical model (3)-(4), ϵ and η are the "location" and "shape" parameters respectively. Therefore, in most cases, it is sufficient to assume that $\eta > 0$. Under this assumption, we have $\eta_+ = \eta$ and $\eta_- \equiv 0$. Consequently,

$$EA = E\eta\,[a_0, b_0]\,,$$

and

$$\begin{aligned} Var(A) &= E\epsilon^2 + \frac{1}{2}(a_0^2 + b_0^2)Var(\eta) + (a_0 + b_0)E\epsilon\eta \\ &= Var(\epsilon) + \frac{1}{2}(a_0^2 + b_0^2)Var(\eta) + (a_0 + b_0)Cov(\epsilon, \eta), \end{aligned}$$

with $E\epsilon = 0$.

3. The minimum contrast estimator.

3.1. *Definitions.* We study the minimum contrast estimator (MCE) of the Normal hierarchical random interval (3)-(4), as well as its asymptotic properties. Since $d = 1$, from now on we let $\mathcal{K}$ be the space of all non-empty compact subsets in $\mathbb{R}$ restrictively, and let $\mathcal{F}$ be the Borel σ-algebra on $\mathcal{K}$ induced by the Hausdorff metric. Let $\mathcal{K}_{\mathcal{C}}$ denote the space of all non-empty compact convex subsets, i.e., bounded closed intervals, in $\mathbb{R}$. As mentioned in the previous section, a random interval X is a Borel measurable function from a probability space $(\Omega, \mathcal{L}, P)$ to $(\mathcal{K}, \mathcal{F})$ such that $X \in \mathcal{K}_{\mathcal{C}}$ almost surely.

Throughout this section, we assume observing a sample of i.i.d. random intervals $X(n) = \{X_1, X_2, \cdots, X_n\}$. Let $\boldsymbol{\theta}$ denote a $p \times 1$ vector containing all the parameters in the model, which takes on a value from a parameter space $\Theta \subset \mathbb{R}^p$. Here p is the number of parameters. Let $\boldsymbol{\theta}_0$ denote the true value of the parameter vector. Denote by $T_{\boldsymbol{\theta}}(a, b)$ the hitting function of X_i with respect to $\boldsymbol{\theta}, \forall [a, b] \in \mathcal{K}_{\mathcal{C}}$.

In order to introduce the MCE, we will need some extra notations. Let $\mathbf{X}$ be a basic set and $\mathcal{A}$ be a σ-field over it. Let $\mathcal{B}$ denote a family of probability measures on ($\mathbf{X}$,$\mathcal{A}$) and τ be a mapping from $\mathcal{B}$ to some topologial space T. $\tau(P)$ denotes the parameter value pertaining to P, $\forall P \in \mathcal{B}$. The classical definition of MCE given in Pfanzagl (1969) is quoted below.

DEFINITION 1. $[Pfanzagl(1969)]$ *A family of $\mathcal{A}$-measurable functions $f_t : \boldsymbol{X} \to \mathbb{R}, t \in T$ is a family of contrast functions if*

$$E_P\,[f_t] < \infty, \tag{10}$$

$\forall t \in T, \forall P \in \mathcal{B}$, *and*

$$E_P\left[f_{\tau(P)}\right] < E_P\left[f_t\right], \tag{11}$$

$\forall t \in T, \forall P \in \mathcal{B}, t \neq \tau(P)$.

In other words, a contrast function is a measurable function of the random variable(s) whose expected value reaches its minimum under the probability measure that generates the random variable(s). From the view of probability, with the true parameters, a contract function tends to have a smaller value than with other parameters. The contrast function is an essential concept in classical statistics. For example, the negative log likelihood (and the negative log density when there is one single observation) is a contrast function, minimizing which leads to the well-known maximum likelihood estimation (MLE).

Adopting some notations from Pfanzagl (1969), we let $\mathcal{B}$ denote a family of probability measures on $(\mathcal{K}_{\mathcal{C}}, \mathcal{F})$ and τ be a mapping from $\mathcal{B}$ to some topologial space T. Similarly, $\tau(P)$ denotes the parameter value pertaining to P, $\forall P \in \mathcal{B}$. We modify the notion of MCE in Heinrich (1993) according to the scenario of random intervals, and give our definition of contrast function below. And then the MCE is defined as the minimizer of the contrast function.

Definition 2. *A family of $\mathcal{F}^n$-measurable functions $M(X(n);\boldsymbol{\theta})$: $\mathcal{K}_{\mathcal{C}}^n \to [-\infty, +\infty]$, $n \in \mathbb{N}$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is a family of contrast functions for $\mathcal{B}$, if there exists a function $N(\cdot,\cdot)$: $\boldsymbol{\Theta} \times \boldsymbol{\Theta} \to \mathbb{R}$ such that*

$$P_{\boldsymbol{\theta}}(\left\{\omega : \lim_{n\to\infty} M(X(n);\boldsymbol{\zeta}) = N(\boldsymbol{\theta},\boldsymbol{\zeta})\right\}) = 1, \ \forall \ \boldsymbol{\theta},\boldsymbol{\zeta} \in \boldsymbol{\Theta}, \tag{12}$$

and

$$N(\boldsymbol{\theta},\boldsymbol{\theta}) < N(\boldsymbol{\theta},\boldsymbol{\zeta}) \ \forall \ \boldsymbol{\theta},\boldsymbol{\zeta} \in \boldsymbol{\Theta}, \ \boldsymbol{\theta} \neq \boldsymbol{\zeta}. \tag{13}$$

Definition 3. *A $\mathcal{F}^n$-measurable function $\hat{\boldsymbol{\theta}}_n$: $\mathcal{K}_{\mathcal{C}}^n \to \tau(\mathcal{B})$, which depends on $X(n)$ only, is called a minimum contrast estimator (MCE) if*

$$M(X(n);\hat{\boldsymbol{\theta}}_n) = \inf\left\{M(X(n);\boldsymbol{\theta}) : \boldsymbol{\theta} \in \tau(\mathcal{B})\right\}. \tag{14}$$

3.2. *Theoretical results.* We make the following assumptions to present the theoretical results in this section.

Assumption 1. Θ *is compact, and* $\boldsymbol{\theta}_0$ *is an interior point of* Θ.

Assumption 2. *The model is identifiable.*

Assumption 3. $T_{\boldsymbol{\theta}}(\cdot,\cdot)$ *is continuous with respect to* $\boldsymbol{\theta}$.

Assumption 4. $\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}(\cdot,\cdot)$, $i = 1, \cdot, p$, *exist and are finite on a bounded region* $S^0 \subset \mathbb{R}^2$.

Assumption 5. $\frac{\partial T_{\boldsymbol{\theta}}}{\partial \theta_j}(\cdot,\cdot)$, $\frac{\partial^2 T_{\boldsymbol{\theta}}}{\partial \theta_j \partial \theta_k}(\cdot,\cdot)$, *and* $\frac{\partial^3 T_{\boldsymbol{\theta}}}{\partial \theta_j \partial \theta_k \partial \theta_l}(\cdot,\cdot)$, $i, j, k = 1, \cdots, p$, *exist and are finite on* S^0 *for* $\boldsymbol{\theta} \in \Theta$.

Assumption 4 and 5 are essential to establish the asymptotic normality for the MCE $\hat{\boldsymbol{\theta}}_n$. It is rather mild and can be met by a large class of capacity functionals. For example, if S^0 is closed, then each $T_{\boldsymbol{\theta}_0}$ with continuous up to third order partial derivatives satisfy both assumptions, as a continuous function on a compact region is always bounded. The following theorem gives sufficient conditions under which the minumum contrast estimator $\hat{\boldsymbol{\theta}}_n$ defined above is strongly consistent.

Theorem 1. *Let* $M(X(n); \boldsymbol{\theta})$ *be a contrast function as in Definition 2 and let* $\hat{\boldsymbol{\theta}}_n$ *be the corresponding MCE. Under the hypothesis of Assumption 1 and in addition if* $M(X(n); \boldsymbol{\theta})$ *is equicontinuous w.r.t.* $\boldsymbol{\theta}$ *for all* $X(n), n = 1, 2, \cdots$, *then,*

$$\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0 \quad a.s., \;\; as\ n \to \infty.$$

Let $[a, b] \in \mathcal{K}_{\mathcal{C}}$. Define an empirical estimator $\hat{T}(a, b; X(n))$ for $T(a, b)$ as:

$$\hat{T}(a, b; X(n)) = \frac{\#\left\{X_i : [a, b] \cap X_i \neq \emptyset, i = 1, \cdots, n\right\}}{n}. \tag{15}$$

Extending the contrast function defined in Heinrich (1993) (for parameters in the Boolean model), we construct a family of functions:

$$H(X(n); \boldsymbol{\theta}) = \iint\limits_{S} \left[T_{\boldsymbol{\theta}}(a, b) - \hat{T}(a, b; X(n))\right]^2 W(a, b) \mathrm{d}a \mathrm{d}b, \tag{16}$$

for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $S \subset S^0 \subset \mathbb{R}^2$, and $W(a, b)$ is a weight function on $[a, b]$ satisfying $0 < W(a, b) < C$, $\forall [a, b] \in \mathcal{K}_{\mathcal{C}}$.

We show in the next Proposition that $H(X(n); \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ defined in (16) is a family of contrast functions for $\boldsymbol{\theta}$. This, together with Theorem 1, immediately yields the strong consistency of the associated MCE. This result is summarized in Corollary 1.

Proposition 1. *Consider that Assumption 2 and Assumption 3 are satisfied. Then $H(X(n);\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, as defined in (16), is a family of contrast functions with limiting function*

$$N(\boldsymbol{\theta},\boldsymbol{\zeta}) = \iint_S \left[T_{\boldsymbol{\theta}}(a,b) - T_{\boldsymbol{\zeta}}(a,b)\right]^2 W(a,b)\mathrm{d}a\mathrm{d}b. \tag{17}$$

In addition, $H(X(n);\boldsymbol{\theta})$ is equicontinuous w.r.t. $\boldsymbol{\theta}$.

Corollary 1. *Consider that Assumption 1, Assumption 2, and Assumption 3 are satisfied. Let $H(X(n);\boldsymbol{\theta})$ be defined as in (16), and*

$$\boldsymbol{\theta}_n^H = \arg\min_{\boldsymbol{\theta}\in\Theta} H\left(X(n);\boldsymbol{\theta}\right). \tag{18}$$

Then

$$\boldsymbol{\theta}_n^H \to \boldsymbol{\theta}_0, \text{ a.s.},$$

as $n \to \infty$.

Next, we show the asymptotic normality for $\boldsymbol{\theta}_n^H$. As a preparation, we first prove the following proposition. The central limit theorem for $\boldsymbol{\theta}_n^H$ is then presented afterwards.

Proposition 2. *Assume the conditions of Lemma 1 (in the Appendix). Define*

$$\frac{\partial H}{\partial \boldsymbol{\theta}}\left(X(n);\boldsymbol{\theta}\right) := \left[\frac{\partial H}{\partial \theta_1}\left(X(n);\boldsymbol{\theta}\right), \cdots, \frac{\partial H}{\partial \theta_p}\left(X(n);\boldsymbol{\theta}\right)\right]^T,$$

as the $p\times 1$ gradient vector of $H\left(X(n);\boldsymbol{\theta}\right)$ w.r.t. $\boldsymbol{\theta}$. Then,

$$\sqrt{n}\left[\frac{\partial H}{\partial \boldsymbol{\theta}}\left(X(n);\boldsymbol{\theta}_0\right)\right] \xrightarrow{\mathcal{D}} N\left(0,\Xi\right),$$

where Ξ is the $p\times p$ symmetric matrix with the $(i,j)^{th}$ component

$$\begin{aligned}\Xi(i,j) \;=\; & 4\iiiint_{S\times S}\left\{P\left(X_1\cap[a,b]\neq\emptyset, X_1\cap[c,d]\neq\emptyset\right) - T_{\boldsymbol{\theta}_0}\left(a,b\right)T_{\boldsymbol{\theta}_0}\left(c,d\right)\right\} \\ & \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_i}\left(a,b\right)\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_j}\left(c,d\right)W(a,b)W(c,d)\mathrm{d}a\mathrm{d}b\mathrm{d}c\mathrm{d}d. \end{aligned}\tag{19}$$

Theorem 2. *Let $H(X(n);\boldsymbol{\theta})$ be defined in (16) and $\boldsymbol{\theta}_n^H$ be defined in (18). Assume the conditions of Corollary 1. If additionally Assumption 5 is satisfied, then*

$$\sqrt{n}\left(\boldsymbol{\theta}_n^H - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{D}} N\left(0, C(T_{\boldsymbol{\theta}_0})^{-1}\Xi C(T_{\boldsymbol{\theta}_0})^{-1}\right), \tag{20}$$

where $C(T_{\boldsymbol{\theta}_0}) = 2\iint\limits_S \left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}\right)^T (a,b)W(a,b)\mathrm{d}a\mathrm{d}b$, *and Ξ is defined in (19).*

4. Simulation. We carry out a small simulation to investigate the performance of the MCE introduced in Definition 3. Assume, in the Normal hierarchical model (3)-(4), that

$$\begin{bmatrix}\epsilon\\ \eta\end{bmatrix} \sim \mathrm{BVN}\left(\begin{bmatrix}0\\ \mu\end{bmatrix}, \Sigma = \begin{bmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{bmatrix}\right), \tag{21}$$

and

$$b_0 = a_0 + 1. \tag{22}$$

The bivariate normal distribution conveniently takes care of the variances and covariance of the location parameter ϵ and the shape parameter η. In addition, as seen in (5) in section 2, it is one distribution that makes the "factors" identically distributed as the observed random intervals in a factor model for multiple random intervals. The removal of the freedom of b_0 is for model identifiability purposes; it is seen that the hitting function T_A is defined via ηa_0 and ηb_0 only. For the simulation, we assign the following parameter values:

$$a_0 = 1, \mu = 20, \Sigma = \begin{bmatrix}10 & 1\\ 1 & 10\end{bmatrix}.$$

According to these values, $P(\eta < 0) = 1.2698 \times 10^{-10}$. Therefore the hitting function is approximately:

$$\begin{aligned}
& T_{\boldsymbol{\theta}}([a,b])\\
= \;& P(a - \eta b_0 \le \epsilon \le b - \eta a_0, \eta \ge 0) + P(a - \eta a_0 \le \epsilon \le b - \eta b_0, \eta < 0)\\
< \;& P(a - \eta b_0 \le \epsilon \le b - \eta a_0, \eta \ge 0) + P(\eta < 0)\\
< \;& P(a - \eta b_0 \le \epsilon \le b - \eta a_0, \eta \ge 0) + 10^{-10}\\
\approx \;& P(a - \eta b_0 \le \epsilon \le b - \eta a_0)\\
= \;& P\left(\begin{bmatrix}1 & a_0\\ -1 & -a_0 - 1\end{bmatrix}\begin{bmatrix}\epsilon\\ \eta\end{bmatrix} \le \begin{bmatrix}b\\ -a\end{bmatrix}\right)\\
= \;& \Phi\left(\begin{bmatrix}b\\ -a\end{bmatrix}; D\begin{bmatrix}0\\ \mu\end{bmatrix}, D\Sigma D'\right),
\end{aligned}$$

where $\Phi(\mathbf{x}; \boldsymbol{\mu}, \Omega)$ is the bivariate normal cdf with mean $\boldsymbol{\mu}$ and covariance Ω, and $D = \begin{bmatrix} 1 & a_0 \\ -1 & -a_0 - 1 \end{bmatrix}$. This is another convenience offered by the bivariate normal distribution.

We simulate a random sample of size n from model (3)-(4) with the assigned parameter values, and then compute the MCE's for the model parameters based on the simulated sample. The process is repeated 10 times independently for each n, and we let $n = 100, 200, 300, 400, 500$, successively, to study the consistency and efficiency of the MCE's. The minimization is carried out in Matlab 2011a using the function *fminsearch.m*. Figure 1 shows one random sample of 100 observations generated from the model. We show the average biases and standard errors of the estimates as functions of the sample size in Figure 2 . Here, the average bias and standard error of the estimates of Σ are the L_2 norms of the average bias and standard error matrices, respectively. As expected from Corollary 1 and Theorem 2, both the bias and the standard error reduce to 0 as sample size grows to infinity. The numerical results are summarized in Table 1.

Finally, we point out that the choice of the region of integration S is important. A larger S usually leads to more accurate estimates, but could also result in more computational complexity. We do not investigate this issue in this paper. However, based on our simulation experience, an S that covers most of the points $(a, b) \in \mathbb{R}^2$ such that $[a, b]$ hits some of the observed intervals, is a good choice as a rule of thumb. In our simulation, $E(A) \approx [20, 40]$, by ignoring a small probability $P(\eta < 0)$. Therefore, we choose $S = \{(x - y, x + y) : 20 \leq x \leq 40, 0 \leq y \leq 10\}$, and the estimates are satisfactory.

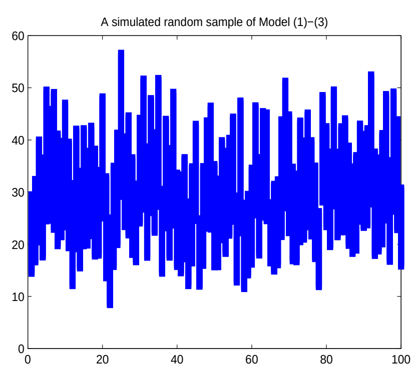

FIG 1. *Plot of a simulated sample from model (3)-(4) with $n = 100$.*

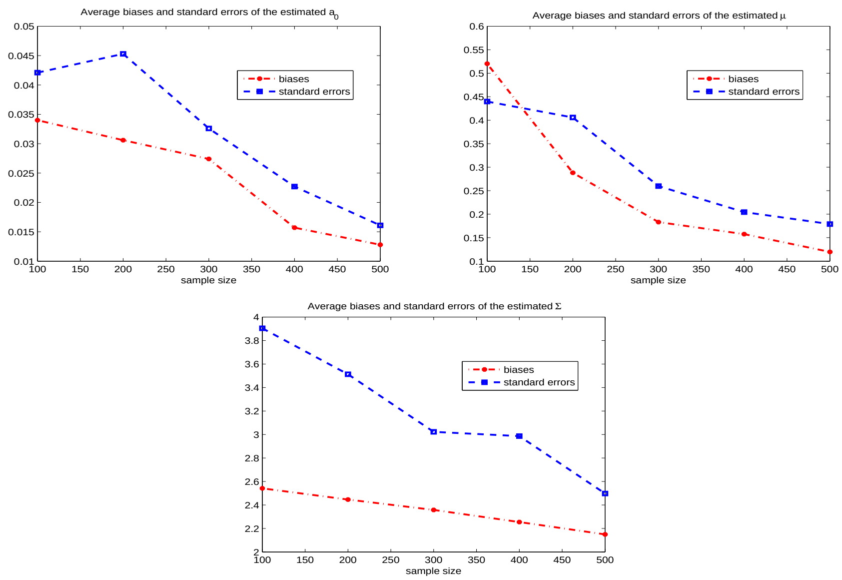

FIG 2. *Average bias and standard error of the MCE's for a_0 (top left), μ (top right), and Σ (bottom), as a function of the sample size n.*

5. A real data application. In this section, we apply our Normal hierarchical model and minimum contrast estimator to analyze the daily temperature range data. We consider two data sets containing ten years of daily minimum and maximum temperatures in January, in Granite Falls, Minnesota (latitude 44.81241, longitude 95.51389) from 1901 to 1910, and from 2001 to 2010, respectively. Each data set, therefore, is constituted of 310 observations of the form: [minimum temperature, maximum temperature] . We obtained these data from the National Weather Service, and all observations are in Fahrenheit. The plot of the data is shown in Figure 3

Same as in the simulation, we assume a bivariate normal distribution for (ϵ, η) and $I_0 = [a_0, a_0 + 1]$ has length 1. The minimum contrast estimates for the model parameters are:

- Data set 1 (1901-1910):

$$\hat{a}_{0,1} = 0.2495, \hat{\mu}_1 = 19.8573, \hat{\Sigma}_1 = \begin{bmatrix} 207.1454 & -44.8547 \\ -44.8547 & 102.5263 \end{bmatrix},$$

- Data set 2 (2001-2010):

$$\hat{a}_{0,2} = 0.2614, \hat{\mu}_2 = 20.4722, \hat{\Sigma}_2 = \begin{bmatrix} 318.9283 & -84.0892 \\ -84.0892 & 68.4783 \end{bmatrix}.$$

TABLE 1
Average biases and standard errors of the MCE's of the model parameters in the simulation study.

n	a_0=1		μ=20		Σ	
	bias	**ste**	**bias**	**ste**	**bias**	**ste**
100	0.0683	0.1289	1.1648	1.7784	4.1166	5.7951
200	0.0387	0.0457	0.4569	0.5924	3.8581	4.0558
300	0.0274	0.0326	0.1831	0.2598	3.0317	3.9042
400	0.0157	0.0227	0.1575	0.2044	2.8210	3.5128
500	0.0128	0.0161	0.1197	0.1790	2.1494	2.4973

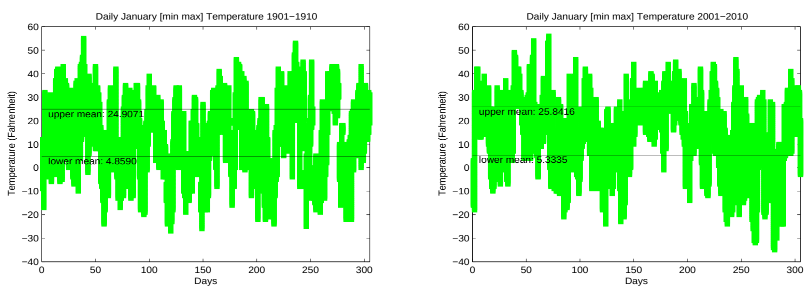

FIG 3. *Plots of daily January temperature range 1901-1910 (left) and 2001-2010 (right). The estimated mean is the interval between the two horizontal black lines, on each plot.*

Denote by A_1 and A_2 respectively the random intervals from which the two data sets are drawn. The estimated mean and variance for A_1 and A_2 are found to be:

$$E(A_1) = [4.8590, 24.9071]\,, Var(A_1) = 221.2313;$$
$$E(A_2) = [5.3335, 25.8416]\,, Var(A_2) = 247.3275.$$

Both mean and variance of the recent data are larger than those of the data 100 years ago. The two estimated means are also shown on the data plots in Figure 3. In addition, the correlation coefficient of (ϵ, η) is -0.3078 for data set 1 and -0.5690 for data set 2, suggesting a negative correlation between the location and the length for the January temperature range data in general. That is, colder days tend to have larger temperature ranges, and, this relationship is stronger in the more recent data.

6. Conclusion. In this paper we introduced a new model of random sets (specifically for random intervals). In many practical situations data are not completely known, or are only known with some margins of error, and it is a very important issue to consider a model which extends normality for ordinary (numerical) data. Our hierarchical normal model extends normality for point-valued random variables, and is quite flexible in the sense that it is well suited for both theoretical investigations and for simulations and real data analysis. To these goals we have defined a minimum contrast estimator for the model parameters, and we have proved its consistency and asymptotic normality. Our main contribution is: distribution-based estimation (as opposed to distance-based, metric space approach). We carry out simulation experiments, and, finally we apply our model to a real data set (daily temperature range data obtained from the National Weather Service). Our approach is suitable for extensions to models in higher dimensions, e.g., a factor model for multiple random intervals, or more general random sets, including possible extensions to spherical random sets.

7. Proofs.

7.1. *Proof of Theorem 1.* Assume by contradiction that $\hat{\boldsymbol{\theta}}_n$ does not converge to $\boldsymbol{\theta}_0$ almost surely. Then, there exists an $\epsilon > 0$ such that

$$P(\left\{\omega : \limsup_{n\to\infty} \left\|\hat{\boldsymbol{\theta}}_n(\omega) - \boldsymbol{\theta}_0\right\| \geq \epsilon\right\}) > 0.$$

Let $F := \left\{\omega : \limsup_{n\to\infty} \left\|\hat{\boldsymbol{\theta}}_n(\omega) - \boldsymbol{\theta}_0\right\| \geq \epsilon\right\}$ and $\Lambda := \Theta \cap \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \epsilon\}$. By the compactness of Λ, for every $\omega \in F$, there exists a convergent subsequence $\left\{\hat{\boldsymbol{\theta}}_{n_i}(\omega)\right\}$ of $\left\{\hat{\boldsymbol{\theta}}_n(\omega)\right\}$ such that

$$\hat{\boldsymbol{\theta}}_{n_i}(\omega) \to \tilde{\boldsymbol{\theta}} \in \Lambda,$$

as $i \to \infty$. Observe that

$$\begin{aligned}
& \liminf_{i\to\infty} M(X(n_i);\boldsymbol{\theta}_0) \\
\geq\ & \liminf_{i\to\infty} M(X(n_i);\hat{\boldsymbol{\theta}}_{n_i}) \\
=\ & \liminf_{i\to\infty} \left\{M(X(n_i);\hat{\boldsymbol{\theta}}_{n_i}) - M(X(n_i);\tilde{\boldsymbol{\theta}}) + M(X(n_i);\tilde{\boldsymbol{\theta}})\right\} \\
\geq\ & \liminf_{i\to\infty} \left\{M(X(n_i);\hat{\boldsymbol{\theta}}_{n_i}) - M(X(n_i);\tilde{\boldsymbol{\theta}})\right\} + \liminf_{i\to\infty} \left\{M(X(n_i);\tilde{\boldsymbol{\theta}})\right\} \\
=\ & \liminf_{i\to\infty} \left\{M(X(n_i);\tilde{\boldsymbol{\theta}})\right\} \qquad (23) \\
=\ & \lim_{n\to\infty} \left\{M(X(n_i);\tilde{\boldsymbol{\theta}})\right\} \\
=\ & N(\tilde{\boldsymbol{\theta}};\boldsymbol{\theta}_0).
\end{aligned}$$

Equation (23) follows from the equicontinuity of $M(X(n);\boldsymbol{\theta})$.

On the other hand,

$$\liminf_{i\to\infty} M(X(n_i);\boldsymbol{\theta}_0) = \lim_{n\to\infty} M(X(n_i);\boldsymbol{\theta}_0) = N(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0).$$

Therefore,

$$P(\left\{\omega : N(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0) \geq N(\tilde{\boldsymbol{\theta}}(\omega),\boldsymbol{\theta}_0)\right\}) > 0, \qquad (24)$$

where $\tilde{\boldsymbol{\theta}} \in \Lambda$ and consequently $\tilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}_0$. But from the assumptions, $N(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0) < N(\tilde{\boldsymbol{\theta}}(\omega),\boldsymbol{\theta}_0), \forall\omega$. This contradicts (24), and therefore completes the proof.

7.2. *Proof of Theorem 2.* From Taylor's Theorem, we have

$$\begin{aligned}
& \frac{\partial H}{\partial\theta_i}\left(X(n);\boldsymbol{\theta}_n^H\right) = 0 \\
=\ & \frac{\partial H}{\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right) + \sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right) \\
& + \frac{1}{2}\left[\sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\frac{\partial}{\partial\theta_j}\right]^2 \frac{\partial H}{\partial\theta_i}\left(X(n);\boldsymbol{\epsilon}_n\right) \\
=\ & \frac{\partial H}{\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right) \\
& + \sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\left[\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}(X(n);\boldsymbol{\theta}_0) + \frac{1}{2}\sum_{l=1}^{p}\left(\theta_{n,l}^H - \theta_{0,l}\right)\frac{\partial^3 H}{\partial\theta_l\partial\theta_j\partial\theta_i}\left(X(n);\boldsymbol{\epsilon}_n\right)\right],
\end{aligned}$$

for $i = 1, \cdots, p$, where $\boldsymbol{\epsilon}_n$ lies between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_n^H$. Writing the above equations in matrix form, we get

$$\frac{\partial H}{\partial \boldsymbol{\theta}}\left(X(n);\boldsymbol{\theta}_0\right) + \left[\frac{\partial^2 H}{\partial \boldsymbol{\theta}^2}\left(X(n);\boldsymbol{\theta}_0\right) + \frac{1}{2}\sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\left(\frac{\partial}{\partial \theta_j}\left(\frac{\partial^2 H}{\partial \boldsymbol{\theta}}\right)\left(X(n);\boldsymbol{\epsilon}_n\right)\right)\right]\left(\boldsymbol{\theta}_n^H - \boldsymbol{\theta}_0\right) = 0.(25)$$

Observe, by taking derivatives under the integral sign, that $\forall i, j$,

$$\begin{aligned}
&\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right)\\
= &\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\iint\limits_S \left[T_{\boldsymbol{\theta}}(a,b) - \hat{T}(a,b;X(n))\right]^2 W(a,b)\mathrm{d}a\mathrm{d}b,\\
= &\frac{\partial}{\partial\theta_j}2\iint\limits_S \left[T_{\boldsymbol{\theta}}(a,b) - \hat{T}(a,b;X(n))\right]\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_i}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b,\\
= &2\iint\limits_S \left[T_{\boldsymbol{\theta}}(a,b) - \hat{T}(a,b;X(n))\right]\frac{\partial^2 T_{\boldsymbol{\theta}_0}}{\partial\theta_j\partial\theta_i}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\\
&+2\iint\limits_S \left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_j}\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_i}\right)(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\\
:= &I + II.
\end{aligned}$$

The first term is

$$\begin{aligned}
I &= 2\iint\limits_S \left(T_{\boldsymbol{\theta}_0}(a,b) - \frac{1}{n}\sum_{k=1}^{n} Y_k(a,b)\right)\frac{\partial^2 T_{\boldsymbol{\theta}_0}}{\partial\theta_j\partial\theta_i}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\\
&= \frac{2}{n}\sum_{k=1}^{n}\iint\limits_S \left[T_{\boldsymbol{\theta}_0}(a,b) - Y_k(a,b)\right]\frac{\partial^2 T_{\boldsymbol{\theta}_0}}{\partial\theta_j\partial\theta_i}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\\
&= o_P(1),
\end{aligned}$$

according to the strong law of large numbers for i.i.d. random variables. Therefore,

$$\frac{\partial^2 H}{\partial\theta_j\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right) = o_P(1) + 2\iint\limits_S \left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_j}\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial\theta_i}\right)(a,b)W(a,b)\mathrm{d}a\mathrm{d}b,$$

$\forall i, j$. In matrix form,

$$\frac{\partial^2 H}{\partial \boldsymbol{\theta}^2}(X(n);\boldsymbol{\theta}_0) = o_P(1) + 2\iint\limits_S \left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \boldsymbol{\theta}}\right)^T (a,b)W(a,b)\mathrm{d}a\mathrm{d}b. \tag{26}$$

Observe again that $\forall j, k, l$,

$$\begin{aligned}
&\left|\frac{\partial^3 H(X(n);\boldsymbol{\epsilon}_n)}{\partial\theta_j\partial\theta_k\partial\theta_l}\right| \\
\leq\ & 2\iint\limits_S \left|\left[T_{\boldsymbol{\epsilon}_n}(a,b) - \hat{T}(a,b;X(n))\right]\frac{\partial^3 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_k\partial\theta_l}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\right| \\
& +2\left|\iint\limits_S \left[\left(\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j}\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_k\partial\theta_l}\right) + \left(\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_k}\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_l}\right) + \left(\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_l}\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_k}\right)\right](a,b)W(a,b)\mathrm{d}a\mathrm{d}b\right| \\
\leq\ & 4\iint\limits_S \left|\frac{\partial^3 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_k\partial\theta_l}(a,b)W(a,b)\mathrm{d}a\mathrm{d}b\right| \\
& +2\left|\iint\limits_S \left[\left(\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j}\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_k\partial\theta_l}\right) + \left(\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_k}\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_l}\right) + \left(\frac{\partial^2 T_{\boldsymbol{\epsilon}_n}}{\partial\theta_j\partial\theta_l}\frac{\partial T_{\boldsymbol{\epsilon}_n}}{\partial\theta_k}\right)\right](a,b)W(a,b)\mathrm{d}a\mathrm{d}b\right| \\
:=\ & C_1(\boldsymbol{\epsilon}_n) \leq C_2,
\end{aligned}$$

$\forall \boldsymbol{\epsilon}_n \in \Theta$, by the compactness of Θ. This, together with the strong consistency of $\boldsymbol{\theta}_n^H$, gives

$$\begin{aligned}
& \frac{1}{2}\sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\left(\frac{\partial}{\partial\theta_j}\left(\frac{\partial^2 H}{\partial\theta_k\partial\theta_l}\right)(X(n);\boldsymbol{\epsilon}_n)\right) \\
=\ & \frac{1}{2}\sum_{j=1}^{p} o_P(1)\frac{\partial^3 H(X(n);\boldsymbol{\epsilon}_n)}{\partial\theta_j\partial\theta_k\partial\theta_l} \\
=\ & o_P(1),
\end{aligned}$$

$\forall k, l$. Equivalently, in matrix form,

$$\frac{1}{2}\sum_{j=1}^{p}\left(\theta_{n,j}^H - \theta_{0,j}\right)\left(\frac{\partial}{\partial\theta_j}\left(\frac{\partial^2 H}{\partial\boldsymbol{\theta}}\right)(X(n);\boldsymbol{\epsilon}_n)\right) = o_P(1). \tag{27}$$

By the multivariate Slutsky's theorem, Proposition 2, together with equation (25), (26), and (27), yields the desired result.

8. Appendix.

8.1. *Proof of Proposition 1.* Notice that $\hat{T}(a,b;X(n))$ is the sample mean of i.i.d. random variables $Y_i:\Omega\to\mathbb{R}$ defined as:

$$Y_i=\begin{cases}1, & \text{if } X_i\cap[a,b]\neq\emptyset,\\ 0, & \text{otherwise.}\end{cases}. \tag{28}$$

Therefore, an application of the strong law of large numbers in the classical case yields:

$$\frac{1}{n}\sum_{i=1}^{n}Y_i\overset{a.s.}{\to}EY_1=P\left(X_1\cap[a,b]\neq\emptyset\right)=T_{\boldsymbol{\theta}_0}\left(a,b\right),\ \text{as } n\to\infty,$$

$\forall a,b:-\infty<a\leq b<\infty$, and assuming $\boldsymbol{\theta}_0$ is the true parameter value. That is,

$$\hat{T}\left(a,b;X(n)\right)\overset{a.s.}{\to}T_{\boldsymbol{\theta}_0}\left(a,b\right),$$

as $n\to\infty$. It follows immediately that

$$\left[\hat{T}(a,b;X(n))-T_{\boldsymbol{\theta}_0}\left(a,b\right)\right]^2W(a,b)\overset{a.s.}{\to}0.$$

Notice that $\forall a,b:-\infty<a\leq b<\infty$, $\left[\hat{T}(a,b;X(n))-T_{\boldsymbol{\theta}_0}\left(a,b\right)\right]^2W(a,b)$ is uniformly bounded by $4C$. By the bounded convergence theorem,

$$\iint_S\left[\hat{T}(a,b;X(n))-T_{\boldsymbol{\theta}_0}\left(a,b\right)\right]^2W(a,b)\mathrm{d}a\mathrm{d}b\overset{a.s.}{\to}\iint_S 0\cdot\mathrm{d}a\mathrm{d}b=0,$$

given any $S\subset\mathbb{R}^2$ with finite Lebesgue measure. This verifies that

$$P_{\boldsymbol{\theta}}\left\{\omega:\lim_{n\to\infty}H\left(X(n);\boldsymbol{\theta}\right)=0\right\}=1. \tag{29}$$

Similarly, we also get

$$P_{\boldsymbol{\theta}}\left\{\omega:\lim_{n\to\infty}H\left(X(n);\boldsymbol{\zeta}\right)=\iint_S\left[T_{\boldsymbol{\theta}}(a,b)-T_{\boldsymbol{\zeta}}(a,b)\right]^2W(a,b)\mathrm{d}a\mathrm{d}b\right\}=1, \tag{30}$$

$\forall\boldsymbol{\theta},\boldsymbol{\zeta}\in\Theta$. Equations (29) and (30) together imply

$$N(\boldsymbol{\theta},\boldsymbol{\zeta})=\iint_S\left[T_{\boldsymbol{\theta}}(a,b)-T_{\boldsymbol{\zeta}}(a,b)\right]^2W(a,b)\mathrm{d}a\mathrm{d}b,\ \boldsymbol{\theta},\boldsymbol{\zeta}\in\Theta. \tag{31}$$

By Assumption 2, $T_{\boldsymbol{\theta}}(a,b) \neq T_{\boldsymbol{\zeta}}(a,b)$, for $\boldsymbol{\theta} \neq \boldsymbol{\zeta}$, except on a Lebesgue set of measure 0. This together with (31) gives

$$N(\boldsymbol{\theta},\boldsymbol{\theta}) < N(\boldsymbol{\theta},\boldsymbol{\zeta}),\ \forall\ \boldsymbol{\theta} \neq \boldsymbol{\zeta},\ \boldsymbol{\theta},\boldsymbol{\zeta} \in \Theta,$$

which proves that $H(X(n);\boldsymbol{\theta})$, $\theta \in \Theta$ is a family of contrast functions. To see the equicontinuity of $H(X(n);\boldsymbol{\theta})$, notice that $\forall \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, we have

$$\begin{aligned}
& |H(X(n);\boldsymbol{\theta}_1) - H(X(n);\boldsymbol{\theta}_2)| \\
= & \ |\iint\limits_S \left(T_{\boldsymbol{\theta}_1}(a,b) - \hat{T}(a,b;X(n))\right)^2 W(a,b)\mathrm{d}a\mathrm{d}b \\
& - \iint\limits_S \left(T_{\boldsymbol{\theta}_2}(a,b) - \hat{T}(a,b;X(n))\right)^2 W(a,b)\mathrm{d}a\mathrm{d}b| \\
= & \ |\iint\limits_S (T_{\boldsymbol{\theta}_1}(a,b) - T_{\boldsymbol{\theta}_2}(a,b)) \left(T_{\boldsymbol{\theta}_1}(r) + T_{\boldsymbol{\theta}_2}(a,b) - 2\hat{T}(a,b;X(n))\right) W(a,b)\mathrm{d}a\mathrm{d}b| \\
\leq & \ 4C \iint\limits_S |T_{\boldsymbol{\theta}_1}(a,b) - T_{\boldsymbol{\theta}_2}(a,b)|\, \mathrm{d}a\mathrm{d}b.
\end{aligned}$$

Then the equicontinuity of $H(X(n);\boldsymbol{\theta})$ follows from the continuity of $T_{\boldsymbol{\theta}}(a,b)$.

8.2. *Lemma 1.* Let $H(X(n);\boldsymbol{\theta})$ be the contrast function defined in (16). Under the hypothesis of Assumption 4,

$$\sqrt{n}\left[\frac{\partial H}{\partial \theta_i}\left(X\left(n\right);\boldsymbol{\theta}_0\right)\right] \xrightarrow{\mathcal{D}} N\left(0,\Delta_i\right),\ \text{as } n \to \infty,$$

for $i = 1, \cdots, p$, where

$$\begin{aligned}
\Delta_i \ = \ & 4 \iiiint\limits_{S\times S} \left\{P\left(X_1 \cap [a,b] \neq \emptyset, X_1 \cap [c,d] \neq \emptyset\right) - T_{\boldsymbol{\theta}_0}\left(a,b\right) T_{\boldsymbol{\theta}_0}\left(c,d\right)\right\} \\
& \times \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}\left(a,b\right) \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}\left(c,d\right) W(a,b)W(c,d)\mathrm{d}a\mathrm{d}b\mathrm{d}c\mathrm{d}d.
\end{aligned}$$

Proof. We will write $\frac{\partial T_{\boldsymbol{\theta}_0}(a,b)}{\partial \theta_i} = T^i_{\boldsymbol{\theta}_0}\left(a,b\right)$ to simplify notations. Exchanging differentiation and integration by the bounded convergence theorem, we get

$$\begin{aligned} &\frac{\partial H}{\partial \theta_i}\left(X\left(n\right);\boldsymbol{\theta}_0\right) \\ = \ & \frac{\partial}{\partial \theta_i}\iint\limits_S \left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-\hat{T}\left(a,b;X(n)\right)\right)^2 W(a,b)\mathrm{d}a\mathrm{d}b \\ = \ & \iint\limits_S \frac{\partial}{\partial \theta_i}\left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-\hat{T}\left(a,b;X(n)\right)\right)^2 W(a,b)\mathrm{d}a\mathrm{d}b \\ = \ & \iint\limits_S 2\left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-\hat{T}\left(a,b;X(n)\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a,b\right) W(a,b)\mathrm{d}a\mathrm{d}b. \end{aligned} \tag{32}$$

Define $Y_i(a,b)$ as in (28). Then,

$$\begin{aligned} (32) \ &= \ \iint\limits_S 2\left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-\frac{1}{n}\sum_{k=1}^{n} Y_k\left(a,b\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a,b\right) W(a,b)\mathrm{d}a\mathrm{d}b \\ &= \ \frac{2}{n}\iint\limits_S \sum_{k=1}^{n}\left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-Y_k\left(a,b\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a,b\right) W(a,b)\mathrm{d}a\mathrm{d}b \\ &= \ \frac{1}{n}\sum_{k=1}^{n} 2\iint\limits_S \left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-Y_k\left(a,b\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a,b\right) W(a,b)\mathrm{d}a\mathrm{d}b \qquad (33) \\ &:= \ \frac{1}{n}\sum_{k=1}^{n} R_k. \end{aligned}$$

Notice that R_k's are i.i.d. random variables: $\Omega \to \mathbb{R}$.

Let $\{\Delta s_1, \Delta s_2, \cdots, \Delta s_m\}$ be a partition of S, and (a_j, b_j) be any point in Δs_j, $j = 1, \cdots, m$. Let $\lambda = \max_{1\le j\le m}\{\mathrm{diam}\Delta s_j\}$. Denote by $\Delta\sigma_j$ the area of Δs_j. By the definition of the double integral,

$$\begin{aligned} R_k \ &= \ 2\iint\limits_S \left(T_{\boldsymbol{\theta}_0}\left(a,b\right)-Y_k\left(a,b\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a,b\right) W(a,b)\mathrm{d}a\mathrm{d}b \\ &= \ \lim_{\lambda\to 0}\left\{\sum_{j=1}^{m}\left(T_{\boldsymbol{\theta}_0}\left(a_j,b_j\right)-Y_k\left(a_j,b_j\right)\right) T^i_{\boldsymbol{\theta}_0}\left(a_j,b_j\right) W(a_j,b_j)\Delta\sigma_j\right\}. \end{aligned}$$

Therefore, by the Lebesgue dominated convergence theorem,

$$
\begin{aligned}
& ER_k \\
= \;& 2E \lim_{\lambda \to 0} \left\{ \sum_{j=1}^{m} \left(T_{\boldsymbol{\theta}_0} (a_j, b_j) - Y_k (a_j, b_j) \right) T^i_{\boldsymbol{\theta}_0} (a_j, b_j) \, W(a_j, b_j) \Delta\sigma_j \right\} \\
= \;& 2 \lim_{\lambda \to 0} \left\{ \sum_{j=1}^{m} \left[E \left(T_{\boldsymbol{\theta}_0} (a_j, b_j) - Y_k (a_j, b_j) \right) \right] T^i_{\boldsymbol{\theta}_0} (a_j, b_j) \, W(a_j, b_j) \Delta\sigma_j \right\} \\
= \;& 2 \lim_{\lambda \to 0} \left\{ \sum_{j=1}^{m} 0 \right\} = 0.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
& Var(R_k) = ER_k^2 \\
= \;& 4E\left\{\lim_{\lambda\to 0}\left\{\sum_{j=1}^{m}\left(T_{\boldsymbol{\theta}_0}\left(a_j,b_j\right)-Y_k\left(a_j,b_j\right)\right)T^i_{\boldsymbol{\theta}_0}\left(a_j,b_j\right)W(a_j,b_j)\Delta\sigma_j\right\}\right\}^2 \\
= \;& 4E\lim_{\lambda_1\to 0}\lim_{\lambda_2\to 0}\left\{\sum_{j_1=1}^{m_1}\left(T_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)-Y_k\left(a_{j_1},b_{j_1}\right)\right)T^i_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)W(a_{j_1},b_{j_1})\Delta\sigma_{j_1}\right\} \\
& \left\{\sum_{j_2=1}^{m_2}\left(T_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)-Y_k\left(a_{j_2},b_{j_2}\right)\right)T^i_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)W(a_{j_2},b_{j_2})\Delta\sigma_{j_2}\right\} \\
= \;& 4E\lim_{\lambda_1\to 0}\lim_{\lambda_2\to 0}\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}\left(T_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)-Y_k\left(a_{j_1},b_{j_1}\right)\right)\left(T_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)-Y_k\left(a_{j_2},b_{j_2}\right)\right) \\
& T^i_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)T^i_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)W(a_{j_1},b_{j_1})W(a_{j_2},b_{j_2})\Delta\sigma_{j_1}\Delta\sigma_{j_2} \\
= \;& 4\lim_{\lambda_1\to 0}\lim_{\lambda_2\to 0}\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}E\left(T_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)-Y_k\left(a_{j_1},b_{j_1}\right)\right)\left(T_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)-Y_k\left(a_{j_2},b_{j_2}\right)\right) \\
& T^i_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)T^i_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)W(a_{j_1},b_{j_1})W(a_{j_2},b_{j_2})\Delta\sigma_{j_1}\Delta\sigma_{j_2} \\
= \;& 4\lim_{\lambda_1\to 0}\lim_{\lambda_2\to 0}\sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2}Cov\left(Y_k\left(a_{j_1},b_{j_1}\right),Y_k\left(a_{j_2},b_{j_2}\right)\right) \\
& T^i_{\boldsymbol{\theta}_0}\left(a_{j_1},b_{j_1}\right)T^i_{\boldsymbol{\theta}_0}\left(a_{j_2},b_{j_2}\right)W(a_{j_1},b_{j_1})W(a_{j_2},b_{j_2})\Delta\sigma_{j_1}\Delta\sigma_{j_2} \\
= \;& 4\iiiint\limits_{S\times S}Cov\left(Y_k\left(a,b\right),Y_k\left(c,d\right)\right)T^i_{\boldsymbol{\theta}_0}\left(a,b\right)T^i_{\boldsymbol{\theta}_0}\left(c,d\right)W(a,b)W(c,d)\mathrm{d}a\mathrm{d}b\mathrm{d}c\mathrm{d}d \\
= \;& 4\iiiint\limits_{S\times S}\left\{P\left(X_k\cap[a,b]\neq\emptyset,X_k\cap[c,d]\neq\emptyset\right)-T_{\boldsymbol{\theta}_0}\left(a,b\right)T_{\boldsymbol{\theta}_0}\left(c,d\right)\right\} \\
& T^i_{\boldsymbol{\theta}_0}\left(a,b\right)T^i_{\boldsymbol{\theta}_0}\left(c,d\right)W(a,b)W(c,d)\mathrm{d}a\mathrm{d}b\mathrm{d}c\mathrm{d}d.
\end{aligned}
$$

From the central limit theorem for i.i.d. random variables, the desired result follows. □

8.3. *Proof of Proposition 2.* By the Cramér-Wold device, it suffices to prove

$$
\sqrt{n}\sum_{i=1}^{p}\lambda_i\frac{\partial H}{\partial\theta_i}\left(X(n);\boldsymbol{\theta}_0\right)\xrightarrow{\mathcal{D}}N\left(0,\sum_{1\le i,j\le p}\lambda_i\lambda_j\Xi(i,j)\right), \tag{34}
$$

for arbitrary real numbers $\lambda_i, i = 1, \cdots, p$. It is easily seen from (33) in the proof of Lemma 1 that

$$\begin{aligned}
&\sum_{i=1}^{p} \lambda_i \frac{\partial H}{\partial \theta_i} (X(n); \boldsymbol{\theta}_0) \\
= \;& \frac{1}{n} \sum_{k=1}^{n} \left(2 \sum_{i=1}^{p} \lambda_i \iint_S (T_{\boldsymbol{\theta}_0}(a,b) - Y_k(a,b)) \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}(a,b) W(a,b) \mathrm{d}a \mathrm{d}b \right) \\
:= \;& \frac{1}{n} \sum_{k=1}^{n} \left(2 \sum_{i=1}^{p} \lambda_i Q_k^i \right).
\end{aligned}$$

By Lemma 1,

$$E\left(2 \sum_{i=1}^{p} \lambda_i Q_k^i\right) = 2 \sum_{i=1}^{p} \lambda_i \cdot 0 = 0.$$

In view of the central limit theorem for i.i.d. random variables, (34) is reduced to proving

$$Var\left(2 \sum_{i=1}^{p} \lambda_i Q_k^i\right) = \sum_{1 \le i,j \le p} \lambda_i \lambda_j \Xi(i,j). \tag{35}$$

By a similar argument as in Lemma 1, together with some algebraic calculations, we obtain

$$\begin{aligned}
& Var\left(2 \sum_{i=1}^{p} \lambda_i Q_k^i\right) \\
= \;& 4 \sum_{1 \le i,j \le p} \lambda_i \lambda_j Cov\left(Q_k^i, Q_k^j\right) \\
= \;& 4 \sum_{1 \le i,j \le p} \lambda_i \lambda_j E\left(\iint_S (T_{\boldsymbol{\theta}_0}(a,b) - Y_k(a,b)) \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}(a,b) W(a,b) \mathrm{d}a \mathrm{d}b \right) \\
& \left(\iint_S (T_{\boldsymbol{\theta}_0}(a,b) - Y_k(a,b)) \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_j}(a,b) W(a,b) \mathrm{d}a \mathrm{d}b \right) \\
= \;& 4 \sum_{1 \le i,j \le p} \lambda_i \lambda_j \iiiint_{S \times S} \left\{ P(X_1 \cap [a,b] \neq \emptyset, X_1 \cap [c,d] \neq \emptyset) - T_{\boldsymbol{\theta}_0}(a,b) T_{\boldsymbol{\theta}_0}(c,d) \right\} \\
& \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_i}(a,b) \frac{\partial T_{\boldsymbol{\theta}_0}}{\partial \theta_j}(c,d) W(a,b) W(c,d) \mathrm{d}a \mathrm{d}b \mathrm{d}c \mathrm{d}d.
\end{aligned}$$

This validates (35), and hence finishes the proof.

References.

[1] Artstein, Z. and Vitale, R.A. (1975). A strong law of large numbers for random compact sets. *Annals of Probability* **5** 879–882.

[2] Aumann, R.J. (1965). Integrals and set-valued functions. *Journal of Mathematical Analysis and Applications* **12** 1–12.

[3] Billard, L. and Diday, E. (2003). From the statistics of data to the statistics of knowledge: symbolic data analysis. *Journal of the American Statistical Association* **98** 462, review article.

[4] Carvalho, F.A.T., Neto, E.A.L., and Tenorio, C.P. (2004). A New Method to Fit a Linear Regression Model for Interval-Valued Data. *Lecture Notes in Computer Science* **3238** 295–306.

[5] Diamond, P. (1990). Least squares fitting of compact set-valued data. *Journal of Mathematical Analysis and Applications* **147** 531–544.

[6] Gil, M.A., Lubiano, M.A., Montenegro, M., and Lopez, M.T. (2002). Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika* **56** 97–111.

[7] Heinrich, L. (1993). Asymptotic properties of minimum contrast estimators for parameters of Boolean models. *Metrika* **40** 69–74.

[8] Hörmander, L. (1954). Sur la fonction d'appui des ensembles convexes dans un espace localement convexe. *arkiv for mat.* **3** 181–186.

[9] Kendall, D.G. (1974). Foundations of a theory of random sets. *In Stochastic Geometry* eds. Harding, E.F. and Kendall, D.G. John Wiley & Sons, New York.

[10] Körner, R. (1995). A Variance of Compact Convex Random Sets. Working paper, Institut fr Stochastik, Bernhard-von-Cotta-Str. 2 09599 Freiberg.

[11] Körner, R. (1997). On the variance of fuzzy random variables. *Fuzzy Sets and Systems* **92** 83–93.

[12] Körner, R. and Näther, W (1998). Linear regression with random fuzzy variables: extended classical estimates, best linear estimates, least squares estimates. *Information Sciences* **109** 95–118.

[13] Lyashenko, N.N. (1979). On limit theorems for sums of independent compact random subsets in the Euclidean space. Translated from *Zapiski Nanchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta* **85** 113–128.

[14] Lyashenko, N.N. (1983). Statistics of random compacts in Euclidean space. Translated from *Zapiski Nanchnykh Seminarov Leningradskogo Otdeleniya Matematicheskogo Instituta* **98**, 115–139.

[15] Matheron, G. (1967). *Elements pour une theorie des milieux poreux*. Masson, Paris.

[16] Matheron, G. (1975). *Random Sets and Integral Geometry*. John Wiley & Sons, New York.

[17] Molchanov, I. (2005). *Theory of Random Sets*. Springer-Verlag, London.

[18] Norberg, T. (1984). Convergence and existence of random set distributions. *The Annals of Probability* **12** 3, 726–732.

[19] Pfanzagl, J. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14**, 249–272.

[20] Puri, M.L., Ralescu, D.A., and Ralescu, S.S. (1986). Gaussian random sets in Banach space. *Theory of Probablity and its Applications* **31** 526–529.

[21] Rådström, H. (1952). An embedding theorem for spaces of convex sets. *Proc. Amer. Math. Soc.* **3** 165–169.

[22] Schneider, R. (1993). *Convex Bodies, the Brunn-Minkowski Theory.* Cambridge University Press, Cambridge.

Mathematics & Statistics
Utah State University
3900 Old Main Hill
Logan, Utah 84322-3900
E-mail: yan.sun@usu.edu

Department of Mathematical Sciences
University of Cincinnati
4199 French Hall
2815 Commons Way
Cincinnati, Ohio 45221-0025
E-mail: dan.ralescu@uc.edu