SOME MEASURE-THEORETIC PROPERTIES OF U-STATISTICS APPLIED IN STATISTICAL PHYSICS

Abstract. This paper investigates the relationship between various measure-theoretic properties of U-statistics with fixed sample size $N$ and the same properties of their kernels. Specifically, the random variables are replaced with elements in some measure space $(\Lambda; dx)$, the resultant real-valued functions on $\Lambda^N$ being called generalized $N$-means. It is shown that a.e. convergence of sequences, measurability, essential boundedness and, under certain conditions, integrability with respect to probability measures of generalized $N$-means and their kernels are equivalent. These results are crucial for the solution of the inverse problem in classical statistical mechanics in the canonical formulation.

1. Introduction

Let $(\Lambda; dx)$ be a complete $\sigma$-finite measure space with non-zero measure $dx$, and let $d^k x$ be the completion of the product measure $dx^\otimes k$ on $\Lambda^k$ for $k \in \mathbb{N}$. If $1 \leq m \leq N$ are integers, and $u : \Lambda^m \to \mathbb{R}$ is a function, the generalized $N$-mean of order $m$ with kernel $u$ is defined in this paper as

\begin{equation}
(G_{N,m} u)(x_1, ..., x_N) = \binom{N}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq N} u(x_{i_1}, ..., x_{i_m}).
\end{equation}

In the following, we investigate whether various measure-theoretic properties of the kernels (such as a.e. convergence of sequences, measurability, essential boundedness, and integrability with respect to probability measures) can be deduced from the analogous properties of the generalized means, and vice versa. Such questions arise in proving the inverse conjecture of statistical physics in the canonical formulation \cite{1, 8}. The inverse hypothesis states that there is a unique kernel $u$ such that the potential of the form (1.1) (traditionally, without the scaling factor) produces a given $m$-particle density. Technically, the kernel $u$ is sought as a minimizer of a relative entropy functional. (Strictly speaking, a maximizer of its negative.) In the canonical formulation, when $m \geq 2$, the a.e. convergence and integrability are first proved for a sequence

\textit{2010 Mathematics Subject Classification.} 28A20, 28A25, 28A35, 62H12, 82B21.

\textit{Key words and phrases.} U-statistics, measurability, a.e. convergence, integrability, inverse problem in statistical mechanics.
of generalized $N$-means and its a.e. limit, respectively [8]. Hence the need for the above mentioned equivalence.

When $m = 1$, the inverse conjecture lies at the foundation of density functional theory (DFT) of inhomogeneous fluids [2]. Its validity for $m \geq 2$, combined with the DFT approach, leads to new results in liquid state theory. (This will be explored in a later paper.) The inverse conjecture for $m \geq 2$ is also customarily assumed in coarse-grained modeling [9,7].

Even though the questions addressed in this paper originated in statistical physics, the results may have broader value. Incidentally, if $(x_1, \ldots, x_N)$ are replaced with random variables, the generalized $N$-mean in (1.1) becomes a U-statistic. U-statistics were introduced by Hoeffding as unbiased estimators of regular functionals [4]. Since then, they have been extensively studied, and numerous applications have been found for them [5,6]. In common statistical usage, the kernel $u$ is given, and the limit properties of U-statistics are studied as sampling size $N$ goes to infinity. These issues are different from the ones dealt with in the inverse problem. Nevertheless, it should be expected that some applications require a measure-theoretic setting in which the properties of U-statistics provided here are useful.

The paper is organized as follows. In Section 2 it is proved that a sequence of generalized $N$-means converges a.e. on $\Lambda^N$ if and only if the corresponding sequence of their kernels converges a.e. on $\Lambda^m$ (Theorem 2.3). Despite its apparent simplicity, the ”only if” part of this statement is not easy to verify. The difficulty lies in the fact that convergence holds only a.e. on $\Lambda^N$. This can be illustrated on a simple example. Suppose that $G_{N,1} u_n \to U$ everywhere on $\Lambda^N$. Then, for every $x \in \Lambda$, $u_n(x) \to U(x, \ldots, x)$. However, the same approach cannot be used if the convergence holds only a.e. because the diagonal may be (and often is) a set of measure zero.

The equivalence of measurability and essential boundedness of generalized means and their kernels is established in Theorem 3.1. Section 4 is concerned with integrability issues. The general problem is as follows. A symmetric probability density $P$ on $\Lambda^N$ induces a marginal symmetric probability density $\rho^{(m)}$ on $\Lambda^m$ upon integrating $P$ with respect to any set of $N - m$ variables. If $1 \leq m < N$ and $1 \leq r < \infty$, is it true that a generalized $N$-mean of order $m$ is in $L^r(\Lambda^N; P d^N x)$ if and only if its kernel is in $L^r(\Lambda^m; \rho^{(m)} d^m x)$? While the ”if” part of this question is easy to verify, the ”only if” part does not hold in general (Example 4.1). However, an extra condition on $P$ given in Theorem 4.3
ensures that the answer to the above question is positive. This condition holds in some arbitrarily small perturbations in $L^1(\Lambda^N; d^N x)$ (and in $L^\infty(\Lambda^N; d^N x)$, if measure $dx$ is finite, and $P$ is essentially bounded) of any symmetric probability density (Theorem 4.5).

We will finish this section by introducing some terminology that will be used throughout the paper.

Subsets of $d^k x$ measure zero will be called null sets, and their complements co-null. The wording "a.e.," "null set," "co-null set" will always be understood relative to the measure $d^k x$, with $k$ obvious from the context, and the same is true regarding the measurability of functions. If a set $E \subset \Lambda^k$ is measurable, its measure will be denoted $|E|$.

We will also use the following definitions. Let $1 \leq m < N$ be integers. Then for any $(x_{m+1}, ..., x_N) \in \Lambda^{N-m}$ and any $E \subset \Lambda^N$, the $(x_{m+1}, ..., x_N)$-section of $E$ is

\begin{equation}
E_{x_{m+1}, ..., x_N} := \{(x_1, ..., x_m) \in \Lambda^m : (x_1, ..., x_N) \in E\}.
\end{equation}

To shorten the presentation, it is useful to define the operator $\hat{B}_{N,m}$, transforming a set $E_m \subset \Lambda^m$ into a set $\hat{B}_{N,m} E_m \subset \Lambda^N$:

\begin{equation}
\hat{B}_{N,m} E_m := \{(x_1, ..., x) \in \Lambda^N : (x_{i_1}, ..., x_{i_m}) \in E_m \ \forall \ 1 \leq i_1 < \cdots < i_m \leq N\}.
\end{equation}

Finally, for ease of future argument, we need to extend the definition of $G_{N,m} u$ to the case where $0 = m \leq N$, and $u \equiv c \in \mathbb{R}$. In this case, we define $G_{N,0} u = c$.

It should be emphasized that, with the exception of Section 4, the kernel $u$ in (1.1) is not assumed to be symmetric, as is customarily done for U-statistics.

2. Almost everywhere convergence

In this section we address the question of whether the a.e. convergence of a sequence of generalized means $G_{N,m} u_n$ implies the a.e. convergence of their kernels $u_n$. In particular, if this is true, then the a.e. limit of a sequence of generalized $N$-means is a generalized $N$-mean.

The following two simple lemmas will be crucial for the development of all subsequent arguments.

**Lemma 2.1.** Let $1 \leq k \leq N$ be integers, and let $T_k$ be a co-null subset of $\Lambda^k$. Then, the set $T_N = \hat{B}_{N,k} T_k$ is co-null in $\Lambda^N$. 

Proof.  
\begin{equation}
T_N = \bigcap_{1 \leq i_1 < \cdots < i_k \leq N} T_{i_1, \ldots, i_k},
\end{equation}
where \( T_{i_1, \ldots, i_k} := \{ (x_1, \ldots, x_N) \in \Lambda^N : (x_{i_1}, \ldots, x_{i_k}) \in T_k \} \). Therefore,
\begin{equation}
|\Lambda^N \setminus T_N| \leq \sum_{1 \leq i_1 < \cdots < i_k \leq N} |\Lambda^N \setminus T_{i_1, \ldots, i_k}| = \binom{N}{k} \left( (\Lambda^k \setminus T_k) \otimes \Lambda^{N-k} \right) = 0.
\end{equation}
\hfill \square

Lemma 2.2. For any integers \( 0 \leq k \leq m \leq N \), \( G_{N,m} G_{m,k} = G_{N,k} \).

Proof. If \( 0 = k \leq m \leq N \), and \( u \equiv c \in \mathbb{R} \) then, \( G_{N,0}u = c \) by definition. On the other hand, \( G_{N,m} G_{m,0}u = G_{N,m}c = c \).

Suppose that \( 1 \leq k \leq m \leq N \), and let \( u : \Lambda^k \to \mathbb{R} \) be any function. By a simple combinatorial argument
\begin{equation}
(G_{N,m} G_{m,k}u)(x_1, \ldots, x_N) = \binom{N}{m}^{-1} \binom{m}{k}^{-1} \binom{N-k}{m-k} \sum_{1 \leq i_1 < \cdots < i_k \leq N} u(x_{i_1}, \ldots, x_{i_k}) = \binom{N}{k}^{-1} \sum_{1 \leq i_1 < \cdots < i_k \leq N} u(x_{i_1}, \ldots, x_{i_k}) = G_{N,k}(x_1, \ldots, x_N).
\end{equation}
\hfill \square

The first equality follows because for every \( 1 \leq i_1 < \cdots < i_k \leq N \), the term \( u(x_{i_1}, \ldots, x_{i_k}) \) appears exactly \( \binom{N-k}{m-k} \) times.

Theorem 2.3. Let \( 1 \leq m \leq N \) be integers, and let \((u_n)\) be a sequence of finite functions on \( \Lambda^m \). Then the following is true.

There is a finite function \( U \) on \( \Lambda^N \) such that \( G_{N,m}u_n \to U \) a.e. if and only if there is a finite function \( u \) on \( \Lambda^m \) such that \( u_n \to u \) a.e.

Remark 2.1. The statement is still true with \( U, u \) and \( u_n \) are a.e. finite. However, we chose them to be everywhere finite to avoid cluttering the proof with non-essential details.

Proof of Theorem 2.3. Since there is nothing to prove when \( m = N \), we will assume that \( 1 \leq m < N \). Suppose first that there exists a finite \( u \) such that \( u_n \to u \) on some co-null set \( E_m \subset \Lambda^m \). Then the set \( E_N = B_{N,m} E_m \subset \Lambda^N \) is co-null by Lemma 2.1. Moreover, \( G_{N,m}u_n \to G_{N,m}u \) on \( E_N \).

Conversely, suppose that there exists a finite \( U \) such that \( G_{N,m}u_n \to U \) on some co-null set \( E \subset \Lambda^N \).
Case 1: $m = 1$. Let us fix $(\tilde{x}_2, ..., \tilde{x}_N) \in \Lambda^{N-1}$ such that $|\Lambda \setminus E_{\tilde{x}_2, ..., \tilde{x}_N}| = 0$. (This is possible because the a.e. section of a co-null set is co-null by the Fubini-Tonelli theorem [3, Theorem 2.39].) By the definition of the set $E$, for every $x \in E_{\tilde{x}_2, ..., \tilde{x}_N}$:

$$
(2.4) \quad u_n(x) + \sum_{j=2}^{N} u_n(\tilde{x}_j) \rightarrow NU(x, \tilde{x}_2, ..., \tilde{x}_N).
$$

Thus, for any $(y_1, ..., y_N) \in E_{\tilde{x}_2, ..., \tilde{x}_N}^N$:

$$
(2.5) \quad \sum_{i=1}^{N} u_n(y_i) + N \sum_{j=2}^{N} u_n(\tilde{x}_j) \rightarrow N \sum_{i=1}^{N} U(y_i, \tilde{x}_2, ..., \tilde{x}_N),
$$

where we have summed (2.4) over $i$ after replacing $x$ with $y_i$. Let us fix $(y_1, ..., y_N) \in E_{\tilde{x}_2, ..., \tilde{x}_N} \cap E$. Then, (2.5) holds together with $\sum_{i=1}^{N} u_n(y_i) \rightarrow NU(y_1, ..., y_N)$. Thus,

$$
(2.6) \quad \sum_{i=2}^{N} u_n(\tilde{x}_i) \rightarrow \sum_{i=1}^{N} U(y_i, \tilde{x}_2, ..., \tilde{x}_N) - U(y_1, ..., y_N) =: C.
$$

Using this result in (2.4), we finally obtain:

$$
(2.7) \quad u_n(x) \rightarrow NU(x, \tilde{x}_2, ..., \tilde{x}_N) - C \quad \forall x \in E_{\tilde{x}_2, ..., \tilde{x}_N}.
$$

Case 2: $2 \leq m < N$. The proof for this case will proceed by induction on $m$. Let us define $M := \min(m, N - m)$. Suppose that the ”only if” statement of the theorem is true for $m - 1$. By the Fubini-Tonelli theorem, we can fix $(\tilde{x}_{m+1}, ..., \tilde{x}_N) \in \Lambda^{N-m}$ such that $|\Lambda^m \setminus E_{\tilde{x}_{m+1}, ..., \tilde{x}_N}| = 0$. By the definition of the set $E$:

$$
(2.8) \quad (G_{N,m}u_n)(\cdot, \tilde{x}_{m+1}, ..., \tilde{x}_N) \rightarrow U(\cdot, \tilde{x}_{m+1}, ..., \tilde{x}_N) \quad \text{on } E_{\tilde{x}_{m+1}, ..., \tilde{x}_N}.
$$

Using the functions $v_n^{(m-k)} : \Lambda^{m-k} \rightarrow \mathbb{R}$, $1 \leq k \leq m - 1$, and constants $v_n^{(0)} \in \mathbb{R}$ defined as:

$$
(2.9) \quad v_n^{(m-k)} := \sum_{1 \leq j_1 < ... < j_k \leq N-m} u_n(\cdot, \tilde{x}_{m+j_1}, ..., \tilde{x}_{m+j_k}) \quad \text{if } 1 \leq k \leq m - 1,
$$

$$
(2.9) \quad v_n^{(0)} := \sum_{1 \leq j_1 < ... < j_m \leq N-m} u_n(\tilde{x}_{m+j_1}, ..., \tilde{x}_{m+j_m}).
$$
the left hand side of (2.8) can be rewritten as

\[
\binom{N}{m} (G_{N,m} u_n)(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) = \\
\left[ u_n + \sum_{k=1}^{M} \left( \frac{m}{m-k} \right) G_{m,m-k} v_n(m-k) \right].
\] (2.10)

By Lemma 2.2, \( G_{m,m-k} = G_{m,m-1} G_{m-1,m-k} \), \( 1 \leq k \leq m \). Thus, the right hand side of (2.10) simplifies to: \( u_n + (G_{m,m-1} \omega_n) \), where \( \omega_n : \Lambda^{m-1} \to \mathbb{R} \) is

\[
\omega_n := \sum_{k=1}^{M} \left( \frac{m}{m-k} \right) G_{m-1,m-k} v_n(m-k).
\] (2.11)

Then, in view of (2.8), we obtain that

\[
u_n + G_{m,m-1} \omega_n \to \binom{N}{m} U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) \text{ on } E_{\tilde{x}_{m+1}, \ldots, \tilde{x}_N}.
\] (2.12)

The set \( \Omega = \hat{B}_{N,m} E_{\tilde{x}_{m+1}, \ldots, \tilde{x}_N} \subset \Lambda^N \) is co-null by Lemma 2.1. Moreover, applying the operator \( G_{N,m} \) to both sides of (2.12), gives that on \( \Omega \):

\[
G_{N,m} u_n + G_{N,m} G_{m,m-1} \omega_n = \\
G_{N,m} u_n + G_{N,m-1} \omega_n \to \binom{N}{m} G_{N,m} U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N).
\] (2.13)

(Lemma 2.2 was used in the equality.) Since (2.13) and \( G_{N,m} u_n \to U \) both hold on \( \Omega \cap E \), we further obtain that

\[
G_{N,m-1} \omega_n \to \binom{N}{m} G_{N,m} U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) - U \text{ on } \Omega \cap E.
\] (2.14)

Now, the induction hypothesis implies that there is a finite function \( \omega : \Lambda^{m-1} \to \mathbb{R} \) such that \( \omega_n \to \omega \) a.e. Further, by the "if" statement of the theorem there is a co-null set \( E_m \subset \Lambda^m \) and a finite function \( \phi : \Lambda^m \to \mathbb{R} \) such that \( G_{m,m-1} \omega_n \to \phi \) on \( E_m \). Using this result in (2.12), we finally obtain:

\[
u_n \to \binom{N}{m} U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) - \phi \text{ on } E_m \cap E_{\tilde{x}_{m+1}, \ldots, \tilde{x}_N}.
\] (2.15)

\[\square\]

**Corollary 2.4.** Let \( 1 \leq m \leq N \) be integers, and \( u \) be a finite function on \( \Lambda^m \). Then, \( G_{N,m} u = 0 \) a.e. if and only if \( u = 0 \) a.e. In particular, the linear operator \( G_{N,m} \) is injective.
Proof. It is easy to verify, using Lemma 2.1, that $G_{N,m}u = 0$ a.e. if $u = 0$ a.e. For the converse, let us define a sequence $(v_n)$ by $v_n = u$ if $n$ is odd and $v_n = 0$ if $n$ is even. Since $G_{N,m}v_n \to 0$ a.e., Theorem 2.3 implies that there is a finite function $v$ on $\Lambda^m$ such that $v_n \to v$ a.e. Then, $v = u = 0$ a.e. by the definition of sequence $(v_n)$.

\section{3. Measurability and Essential Boundedness}

**Theorem 3.1.** Let $1 \leq m \leq N$ be integers, and $u : \Lambda^m \to \mathbb{R}$ be a function. Let $U := G_{N,m}u$. Then,

(i) $U$ is measurable if and only if $u$ is measurable.

(ii) $U \in L^\infty(\Lambda^N; d\Lambda x)$ if and only if $u \in L^\infty(\Lambda^m; d\Lambda^m x)$. Moreover, there is a constant $C(N, m)$ such that $|U|_{\infty, d\Lambda^N x} \leq |u|_{\infty, d\Lambda^m x} \leq C(N, m)|U|_{\infty, d\Lambda x}$, with $C(N, 1) = 1$. In particular, $G_{N,m}$ is an isomorphism from $L^\infty(\Lambda^m; d\Lambda^m x)$ onto a closed subspace of $L^\infty(\Lambda^N; d\Lambda x)$.

**Proof.** "In particular" part is the direct consequence of Theorem 2.3 and Corollary 2.4. The conclusion holds trivially for $m = N$, so let us assume that $1 \leq m < N$. Since the proofs of (i) and (ii) are very similar, we will only show (ii). However, it will be clear that (i) is established by a shorter version of the same argument.

Suppose first that $u \in L^\infty(\Lambda^m; d\Lambda^m x)$. Then, $U$ is a finite sum of measurable functions. Namely, $U = \sum_{1 \leq i_1 < \ldots < i_m \leq N} U_{i_1, \ldots, i_m}(x_{i_1}, \ldots, x_{i_m}) := u(x_{i_1}, \ldots, x_{i_m})$. Moreover, the set $E_m \subset \Lambda^m$ on which $|u| \leq |u|_{\infty, d\Lambda^m x}$ is co-null. It follows that $|U| \leq |u|_{\infty, d\Lambda^m x}$ on the co-null set $\hat{B}_{N,m}E_m \subset \Lambda^N$, and so $|U|_{\infty, d\Lambda^N x} \leq |u|_{\infty, d\Lambda^m x}$.

Conversely, suppose that $U \in L^\infty(\Lambda^N; d\Lambda x)$.

**Case 1:** $m = 1$. The Fubini-Tonelli theorem implies that there is $(\tilde{x}_1, \ldots, \tilde{x}_N) \in \Lambda^{N-1}$ such that $U(\cdot, \tilde{x}_1, \ldots, \tilde{x}_N) = u(\cdot) + \sum_{i=2}^N u(\tilde{x}_i) \in L^\infty(\Lambda; dx)$, and so $u \in L^\infty(\Lambda; dx)$.

Next, we will show that $|u|_{\infty, dx} = |U|_{\infty, d\Lambda^N x}$. In view of the "if" part of the theorem, it suffices to prove that $|U|_{\infty, d\Lambda^N x} \geq |u|_{\infty, dx}$.

Let $s := \text{ess sup } u$, and $S := \text{ess sup } U$. For every $\epsilon > 0$, the measure of the set $A_\epsilon := \{x \in \Lambda : u(x) > s - \epsilon\}$ is strictly positive. Since the set $E := \{(x_1, \ldots, x_N) \in \Lambda^N : U(x_1, \ldots, x_N) \leq S\} \subset \Lambda^N$ is co-null, it follows that $|A_\epsilon^N \cap E| > 0$. Moreover, for every $(x_1, \ldots, x_N) \in A_\epsilon^N \cap E$, $S \geq U(x_1, \ldots, x_N) > s - \epsilon$. Thus, $S \geq s$. Similarly, ess inf $U \leq \text{ess inf } u$, and so $|U|_{\infty, d\Lambda^N x} \geq |u|_{\infty, dx}$.

**Case 2:** $2 \leq m < N$. The proof will proceed by induction on $m$. Suppose that the "only if" statement of the theorem holds for $m - 1$. The Fubini-Tonelli theorem implies that there is $(\tilde{x}_{m+1}, \ldots, \tilde{x}_N) \in \Lambda^{N-m}$
such that \( U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) \in L^\infty(\Lambda^m; d^m x) \), with
\[
||U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)||_{\infty, d^m x} \leq ||U||_{\infty, d^N x}.
\]
By the proof of Theorem 2.3 (see (2.8 - 2.12)), there is \( \omega : \Lambda^{m-1} \rightarrow \mathbb{R} \) such that
\[
(3.1)
||U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)||_{\infty, d^m x} \leq ||U||_{\infty, d^N x}.
\]
Applying \( G_{N,m} \) to both sides of (3.2), and using Lemma 2.2 yield:
\[
(3.3)
G_{N,m-1}\omega = \left( \binom{N}{m} \right) G_{N,m}U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) - U.
\]
Therefore, by the "if" statement of the theorem and (3.1), \( G_{N,m-1}\omega \in L^\infty(\Lambda^N; d^N x) \), with
\[
||G_{N,m-1}\omega||_{\infty, d^N x} \leq \left( \binom{N}{m} + 1 \right) ||U||_{\infty, d^N x}.
\]
Then, the induction hypothesis implies that \( \omega \in L^\infty(\Lambda^{m-1}; d^{m-1} x) \), and
\[
(3.5)
||\omega||_{\infty, d^{m-1} x} \leq C(N, m-1) \left( \binom{N}{m} + 1 \right) ||U||_{\infty, d^N x}
\]
for some constant \( C(N, m-1) \). Next, using the "if" part of the theorem again, we infer that \( G_{m,m-1}\omega \in L^\infty(\Lambda^m; d^m x) \), and
\[
||G_{m,m-1}\omega||_{\infty, d^m x} \leq ||\omega||_{\infty, d^{m-1} x}.
\]
This result, together with (3.2), (3.1), and (3.5), yield that \( u \in L^\infty(\Lambda^m; d^m x) \), and
\[
||u||_{\infty, d^m x} \leq \left( \binom{N}{m} \right) ||U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)||_{\infty, d^m x} + ||\omega||_{\infty, d^{m-1} x}
\]
(3.6)
\[
\leq \left[ \binom{N}{m} (1 + C(N, m-1)) + C(N, m-1) \right] ||U||_{\infty, d^N x}.
\]

4. Integrability

Let \( N \geq 2 \) and \( P \) be a symmetric probability density on \( \Lambda^N \). That is, \( P \) is a nonnegative, symmetric function, and \( \int_{\Lambda^N} P = 1 \). For every \( 1 \leq m < N \), the marginal probability density on \( \Lambda^m \) is defined as
\[
(4.1)
\rho^{(m)} := \int_{\Lambda^{N-m}} P(\cdot, x_{m+1}, \ldots, x_N) dx_{m+1} \cdots dx_N \text{ a.e. on } \Lambda^m.
\]
Note that $\rho^{(m)}$ is symmetric a.e. on $\Lambda^m$. In this section we will discuss the relationship between integrability of generalized $N$-means with respect to measure $Pd^N x$ and integrability of their kernels with respect to measure $\rho^{(m)}d^m x$.

It is easy to convince oneself that $u \in L^r(\Lambda^m; \rho^{(m)}d^m x)$ implies that $G_{N,m}u \in L^r(\Lambda^N; Pd^N x)$ for $1 \leq r < \infty$. However, the converse is not obvious and, in fact, is not true in general. This situation is illustrated with the following example.

**Example 4.1.** Consider a $\sigma$-finite measure space $(\Lambda; d\mu)$, where $\Lambda = \mathbb{N}$, and $d\mu$ is the counting measure. Let us define the probability density $P$ on $\Lambda^2$ by the formula:

\begin{equation}
P(i, j) := \begin{cases} 
\frac{1}{(i+j)^2} & \text{if } |i - j| = 1 \\
0 & \text{if } |i - j| \neq 1
\end{cases}
\end{equation}

Then, $P$ is symmetric, and

\begin{equation}
\int_{\Lambda^2} Pd^2 \mu = 2 \sum_{i=1}^{\infty} \frac{1}{(2i+1)^2} < \infty.
\end{equation}

(That $\int_{\Lambda^2} Pd^2 \mu \neq 1$ is immaterial.) For every $i \geq 2$, the marginal probability density $\rho^{(1)}(i)$ is calculated to be

\begin{equation}
\rho^{(1)}(i) = \sum_{j=1}^{\infty} P(i, j) = \frac{1}{(2i+1)^2} + \frac{1}{(2i-1)^2} > \frac{1}{(2i+1)^2}.
\end{equation}

Let us define $u : \Lambda \to \mathbb{R}$ by $u(i) = 2(-1)^i i$. Then, $|u(i) + u(j)| = 2$ whenever $|i - j| = 1$. Thus, using (4.2) and (4.4), we find that

\begin{equation}
\int_{\Lambda^2} P|G_{2,1}u|d^2 \mu = 2 \sum_{i=1}^{\infty} \frac{1}{(2i+1)^2} < \infty,
\end{equation}

but

\begin{equation}
\int_{\Lambda} \rho^{(1)}|u|d\mu > 2 \sum_{i=2}^{\infty} \frac{i}{(2i+1)^2} = \infty.
\end{equation}

In spite of Example 4.1, an extra condition on $P$ ensures that a generalized $N$-mean of order $m$ is in $L^1(\Lambda^N; Pd^N x)$ if and only if its kernel is in $L^1(\Lambda^m; \rho^{(m)}d^m x)$. The generality of this condition is addressed in Theorem 4.5. We begin with two lemmas that will be used in the proof of Theorem 4.3, the main result of this section.

**Lemma 4.1.** Let $1 \leq m \leq N - 1$ be integers, $A \subset \Lambda$ be a subset of positive measure, and $\gamma : A \to (0, \infty)$ be a measurable function. If,

\begin{equation}
\rho^{(m+1)}(x_1, \ldots, x_{m+1}) \geq \gamma(x_{m+1})\rho^{(m)}(x_1, \ldots, x_m)
\end{equation}

then
for a.e. \((x_1, \ldots, x_{m+1}) \in \Lambda^m \otimes A\), then

\[\rho^{(m)}(x_1, \ldots, x_m) \geq \gamma(x_m)\rho^{(m-1)}(x_1, \ldots, x_{m-1})\]

for a.e. \((x_1, \ldots, x_m) \in \Lambda^{m-1} \otimes A\).

**Proof.** Let \(E_{m+1} \subset \Lambda^{m+1}\) be a co-null set such that (4.7) holds on \(E_{m+1} \cap (\Lambda^m \otimes A)\). By the Fubini-Tonelli theorem, there is a co-null set \(E_m \subset \Lambda^m\) such that for every \((x_1, \ldots, x_{m-1}, x_{m+1}) \in E_m\) both sides of (4.7) are integrable functions of \(x_m\), and the section \((E_{m+1})_{x_1,\ldots,x_{m-1},x_{m+1}} \subset \Lambda\) is co-null. If \((x_1, \ldots, x_{m-1}, x_{m+1}) \in E_m \cap (\Lambda^{m-1} \otimes A)\), and \(x_m \in (E_{m+1})_{x_1,\ldots,x_{m-1},x_{m+1}}\), then \((x_1, \ldots, x_{m+1}) \in E_{m+1} \cap (\Lambda^m \otimes A)\), and so inequality (4.7) holds. Thus, integrating both sides of (4.7) with respect to \(x_m\), and subsequently renaming \(m + 1\) with \(m\) yield:

\[\rho^{(m)}(x_1, \ldots, x_m) \geq \gamma(x_m)\rho^{(m-1)}(x_1, \ldots, x_{m-1})\]

for every \((x_1, \ldots, x_m) \in E_m \cap (\Lambda^{m-1} \otimes A)\). \(\square\)

**Lemma 4.2.** Let \(1 \leq m \leq N - 1\) be integers, \(A \subset \Lambda\) be a subset of positive measure, and \(\gamma : A \to (0, \infty)\) be a measurable function. Suppose that

\[P(x_1, \ldots, x_N) \geq \gamma(x_N)\rho^{(N-1)}(x_1, \ldots, x_{N-1})\]

for a.e. \((x_1, \ldots, x_N) \in \Lambda^{N-1} \otimes A\). Then,

\[P(x_1, \ldots, x_N) \geq \gamma(x_N) \cdots \gamma(x_{m+1})\rho^{(m)}(x_1, \ldots, x_m)\]

for a.e. \((x_1, \ldots, x_N) \in \Lambda^m \otimes A^{N-m}\).

**Proof.** Inequality (4.11) clearly holds when \(m = N - 1\). Suppose that it is satisfied for some \(2 \leq m \leq N - 1\). We will show that it then holds for \(m - 1\), and so the lemma will follow by induction.

Using Lemma 4.1, we infer from (4.10) by induction that

\[\rho^{(m)}(x_1, \ldots, x_m) \geq \gamma(x_m)\rho^{(m-1)}(x_1, \ldots, x_{m-1})\]

for a.e. \((x_1, \ldots, x_m) \in \Lambda^{m-1} \otimes A\). Therefore, in view of (4.11),

\[P(x_1, \ldots, x_N) \geq \gamma(x_N) \cdots \gamma(x_m)\rho^{(m-1)}(x_1, \ldots, x_{m-1})\]

for a.e. \((x_1, \ldots, x_N) \in \Lambda^{m-1} \otimes A^{N-m-1}\). \(\square\)

**Theorem 4.3.** Suppose \(N \geq 2\) is an integer, and that for every \(x_N\) in some subset \(A \subset \Lambda\) of positive measure, there is a constant \(\gamma(x_N) > 0\) such that

\[P(\cdot, x_N) \geq \gamma(x_N)\rho^{(N-1)} \ a.e. \ on \ \Lambda^{N-1}\]

Let \(1 \leq m \leq N\) be integers, \(1 \leq r < \infty\), and \(u : \Lambda^m \to \mathbb{R}\) be a function. Define \(U := G_{N,m,u}\). Then,
\[ U \in L^r(\Lambda^N; Pd^N x) \text{ if and only if } u \in L^r(\Lambda^m; \rho^{(m)} d^m x). \] Moreover, there is a constant \( C := C(N, m, r, P) \) such that
\[ ||U||_{r,Pd^N x} \leq ||u||_{r,\rho^{(m)} d^m x} \leq C||U||_{r,Pd^N x}. \] In particular, \( G_{N,m} \) is an isomorphism from \( L^r(\Lambda^m; d^m x) \) onto a closed subspace of \( L^r(\Lambda^N; d^N x) \).

**Remark 4.1.** The condition on \( P \) at the beginning of Theorem 4.3 can be replaced with another, seemingly stronger, but in fact equivalent, assumption. To be specific, we can assume that
\[ (4.15) \quad P \geq \rho^{(N-1)} \otimes \gamma \text{ a.e. on } \Lambda^{N-1} \otimes A, \]
where \( A \subset \Lambda \) is some subset of positive measure, and \( \gamma : A \to (0, \infty) \) is a measurable function.

Indeed, the condition on \( P \) stated in Theorem 4.3 is equivalent to:
\[ \text{ess inf } f(\cdot, x_N) > 0 \text{ for every } x_N \in A, \] where \( f \) is a measurable function on \( \Lambda^N \) defined by
\[ (4.16) \quad f = \begin{cases} P/(\rho^{(N-1)} \otimes 1) & \text{if } \rho^{(N-1)} \otimes 1 > 0, \\ 1 & \text{if } \rho^{(N-1)} \otimes 1 = 0. \end{cases} \]
Moreover, (4.14) holds with \( \gamma(x_N) \) replaced by ess inf \( f(\cdot, x_N) \), a measurable function. However, if \( \gamma \) in (4.14) is \( dx \) measurable, then \( g := P - \rho^{(N-1)} \otimes \gamma \) is \( d^N x \) measurable. Let \( T := \{(x_1, \ldots, x_N) \in \Lambda^N : g(x_1, \ldots, x_N) \geq 0\} \). Arguing by contradiction, it is easy to see that \( |\Lambda^{N-1} \otimes A \setminus T| = 0 \), i.e. (4.15) holds a.e. on \( \Lambda^{N-1} \otimes A \).

**Proof of Theorem 4.3.** The "in particular" part is the direct consequence of Theorem 2.3 and Corollary 2.4. Since there is nothing to prove when \( m = N \), we will assume that \( 1 \leq m < N \).

For the "if" part of the theorem, suppose that \( u \in L^r(\Lambda^m; d^m x) \). Then,
\[ (4.17) \quad \left( \begin{array}{c} N \\ m \end{array} \right)^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq N} \int_{\Lambda^N} |u(x_{i_1}, \ldots, x_{i_m})|^r Pdx_1 \cdots dx_N = \int_{\Lambda^m} |u|^r \rho^{(m)} = ||u||_{r,\rho^{(m)} d^m x}. \]

For the "only if" part, we will use the condition on \( P \), as stated in Remark 4.1. Suppose that \( U \in L^r(\Lambda^N; Pd^N x) \). Since \( \gamma \) is positive on \( A \), there is \( \varepsilon > 0 \) such that the set \( A_\varepsilon := \{x \in A : \gamma(x) > \varepsilon\} \) has positive measure. Thus, in view of Lemma 4.2, for every \( 1 \leq m \leq N - 1 \):
\[ (4.18) \quad P(x_1, \ldots, x_N) \geq \gamma(x_N) \cdots \gamma(x_{m+1}) \rho^{(m)}(x_1, \ldots, x_m) > \varepsilon^{N-m} \rho^{(m)}(x_1, \ldots, x_m) \]
for a.e. \((x_1, \ldots, x_N) \in \Lambda^m \otimes A^{N-m}_\varepsilon\). Note, that \(|A_\varepsilon| < \infty\) because the integration of (4.18) gives \(1 \geq \int_{\Lambda^m \otimes A^{N-m}_\varepsilon} P \geq (\varepsilon|A_\varepsilon|)^{N-m}\).

**Lemma 4.4.** Let \(\alpha := (\varepsilon|A_\varepsilon|)^{-1}\). For every \(1 \leq m \leq N - 1\), the set

\[
T_{N-m} := \{(x_{m+1}, \ldots, x_N) \in A^{N-m}_\varepsilon : U(\cdot, x_{m+1}, \ldots, x_N) \text{ is measurable and}
\]

\[
||U(\cdot, x_{m+1}, \ldots, x_N)||_{r, \rho(m)}^{r} \leq \alpha^{N-m}||U||_{r, \rho_d}^{r}
\]

is not a set of measure zero. In particular, it is not empty.

**Proof.** Suppose that \(|T_{N-m}| = 0\). This means that

\[
||U||_{r, \rho_d}^{r} < \alpha^{m-N} \int_{\Lambda^m} |U(\cdot, x_{m+1}, \ldots, x_N)|^{r} \rho^{(m)} d^m x
\]

for a.e. \((x_{m+1}, \ldots, x_N) \in A^{N-m}_\varepsilon\). Integration over \(A^{N-m}_\varepsilon\) in the last inequality, and (4.18) give:

\[
||U||_{r, \rho_d}^{r} < \int_{\Lambda^m \otimes A^{N-m}_\varepsilon} |U(x_1, \ldots, x_N)|^{r} \rho^{(m)}(x_1, \ldots, x_m) d^N x
\]

\[
< \int_{\Lambda^m \otimes A^{N-m}_\varepsilon} |U|^r \rho_d d^N x \leq ||U||_{r, \rho_d}^{r},
\]

a contradiction. Thus, \(T_{N-m}\) can not be a set of measure zero. \(\Box\)

**Case 1: \(m = 1\).** Let us fix \((\tilde{x}_2, \ldots, \tilde{x}_N) \in T_{N-1}\), a non-empty set by Lemma 4.4. Then,

\[
NU(\cdot, \tilde{x}_2, \ldots, \tilde{x}_N) = u + \sum_{i=2}^{N} u(\tilde{x}_i) := u + \tilde{c}.
\]

By the definition of the set \(T_{N-1}\) given in (4.19),

\[
||U||_{r, \rho_d}^{r} d x \leq \alpha^{N-1} ||U||_{r, \rho_d}^{r} d x.
\]

To get an estimate on \(\tilde{c}\), let us apply \(G_{N,1}\) to both sides of (4.22) with the result \(\tilde{c} = NG_{N,1}U(\cdot, \tilde{x}_2, \ldots, \tilde{x}_N) - U\). Then, the "if" part of the theorem and (4.23) imply that

\[
|\tilde{c}| \leq N||G_{N,1}U(\cdot, \tilde{x}_2, \ldots, \tilde{x}_N)||_{r, \rho_d}^{r} + ||U||_{r, \rho_d}^{r} d x
\]

\[
\leq N||U(\cdot, \tilde{x}_2, \ldots, \tilde{x}_N)||_{r, \rho_d}^{r} d x + ||U||_{r, \rho_d}^{r} d x
\]

\[
\leq \left[ N \alpha^{\frac{N-1}{r}} + 1 \right] ||U||_{r, \rho_d}^{r} d x.
\]
Finally, we infer from (4.22), (4.23), and (4.24) that

\[ \|u\|_{r,\rho(1)}dx \leq N\|U(\cdot, \tilde{x}_2, \ldots, \tilde{x}_N)\|_{r,\rho(1)}dx + |\tilde{c}| \]

(4.25)

\[ \leq \left(2N\alpha^{\frac{N-1}{r}} + 1\right)\|U\|_{r,Pd^N}x. \]

Note that the constant in the round brackets depends on \( P \) through \( \alpha \).

**Case 2:** \( 2 \leq m < N \). Similarly to the proofs for this case in the previous two theorems, we will use induction on \( m \). Suppose that the "only if" statement of the theorem holds for \( m - 1 \). Let us fix \( (\tilde{x}_{m+1}, \ldots, \tilde{x}_N) \in T_{N-m} \), a non-empty set by Lemma 4.4. As was shown previously, (see (2.8 - 2.12) ), there is \( \omega : \Lambda^{m-1} \rightarrow \mathbb{R} \) such that

\[ \left( \begin{array}{c} N \\ m \end{array} \right) U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) = u + G_{m,m-1}\omega. \]

(4.26)

By the definition of the set \( T_{N-m} \) given in (4.19),

\[ \|U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)\|_{r,\rho(m)d^m}x \leq \alpha^{\frac{N-m}{r}}\|U\|_{r,Pd^N}x. \]

(4.27)

An estimate on \( \|G_{m,m-1}\omega\|_{r,\rho(m)d^m}x \) will follow by induction. Applying \( G_{N,m} \) to both sides of (4.26) and using Lemma 2.2 yield:

\[ G_{N,m-1}\omega = \left( \begin{array}{c} N \\ m \end{array} \right) G_{N,m}U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N) - U. \]

(4.28)

The last equation shows that \( G_{N,m-1}\omega \) is measurable. In addition, using the "if" statement of the theorem and (4.27), we estimate:

\[ \|G_{N,m-1}\omega\|_{r,Pd^N}x \leq \left( \begin{array}{c} N \\ m \end{array} \right) \|G_{N,m}U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)\|_{r,Pd^N}x + \|U\|_{r,Pd^N}x \]

\[ \leq \left( \begin{array}{c} N \\ m \end{array} \right) \|U(\cdot, \tilde{x}_{m+1}, \ldots, \tilde{x}_N)\|_{r,\rho(m)d^m}x + \|U\|_{r,Pd^N}x \]

\[ \leq \left[ \left( \begin{array}{c} N \\ m \end{array} \right) \alpha^{\frac{N-m}{r}} + 1 \right] \|U\|_{r,Pd^N}x. \]

(4.29)

The last inequality allows us to conclude from the induction hypothesis that \( \omega \in L^r(\Lambda^{m-1}; \rho^{(m-1)}d^{m-1})x \), and

\[ \|\omega\|_{r,\rho^{(m-1)}d^{m-1}}x \leq \tilde{C} \left[ \left( \begin{array}{c} N \\ m \end{array} \right) \alpha^{\frac{N-m}{r}} + 1 \right] \|U\|_{r,Pd^N}x \]

(4.30)

for some constant \( \tilde{C} = \tilde{C}(N, m - 1, r, P) \). Using the "if" statement of the theorem one more time, we infer that \( G_{m,m-1}\omega \in L^r(\Lambda^m; \rho^{(m)}d^m) \), and
\[ |G_{m,m-1} \omega|_{r, \rho(m) d^m x} \leq |\omega|_{r, \rho(m-1) d^{m-1} x}. \]

This inequality, together with (4.26), (4.27), and (4.30), finally give that \( u \in L^r(\Lambda^m; \rho^{(m)} d^m x) \), and

\[ ||u||_{r, \rho(m) d^m x} \leq \left( \frac{N}{m} \right) \alpha \frac{N-m}{m} (1 + \tilde{C}) + \tilde{C} ||U||_{r, \rho d^N x}. \]

(4.31)

The next theorem shows that any symmetric probability density on \( \Lambda^N \) can be approximated in \( L^1(\Lambda^N; d^N x) \) by an arbitrarily close symmetric probability density satisfying condition (4.14). Moreover, if measure \( dx \) is finite and \( P \in L^\infty(\Lambda^N; d^N x) \), then this approximation is in \( L^\infty(\Lambda^N; d^N x) \).

**Theorem 4.5.** If \( N \geq 2 \) and \( P \) is a symmetric probability density on \( \Lambda^N \), there is a sequence \( (P_n) \) of symmetric probability densities on \( \Lambda^N \) such that \( P_n \) satisfies (4.14), and \( P_n \to P \) in \( L^1(\Lambda^N; d^N x) \). If, in addition, measure \( dx \) is finite, and \( P \) is essentially bounded, then \( P_n \to P \) in \( L^\infty(\Lambda^N; d^N x) \).

**Proof.** It suffices to prove the theorem when measure \( dx \) is finite, and \( P \) is essentially bounded. Indeed, since \( (\Lambda; dx) \) is \( \sigma \)-finite, \( \Lambda = \cup_{n=1}^\infty E_n \), where \( |E_n| < \infty \) \( \forall n \), and \( E_n \subset E_{n+1} \). Then, by dominated convergence, any symmetric probability density \( P \) can be approximated in \( L^1(\Lambda^N; d^N x) \) by a sequence of symmetric probability densities \( P_n \chi_{E_n}/||P_n \chi_{E_n}||_{1,d^N x} \), where \( \chi_{E_n} \) is the characteristic function of the set \( E_n \), and \( P_n = \min\{P, n\} \).

In accordance with the above comment, suppose that \( |\Lambda| < \infty \), and \( P \) is essentially bounded. Then, there is \( c > 0 \) such that \( \rho^{(N-1)} \leq c \) a.e. on \( \Lambda^{N-1} \). If \( Q_n := \max\{P, \frac{1}{n}\} \), then \( P_n := \frac{Q_n}{||Q_n||_{1,d^N x}} \) is a symmetric probability density on \( \Lambda^N \), and \( P_n \to P \) in \( L^1(\Lambda^N; d^N x) \) by dominated convergence. In addition, since \( \frac{1}{n} \leq P - P_n \leq ||P||_{\infty,d^N x} \left( 1 - 1/||Q_n||_{1,d^N x} \right) \) on some co-null set for \( n \) large enough, \( P_n \rightarrow P \) in \( L^\infty(\Lambda^N; d^N x) \).

It remains to check that \( P_n \) satisfies (4.14). For this, we notice that

\[ P_n \geq \frac{1}{n||Q_n||_{1,d^N x}} \geq \frac{1}{n + \Lambda^N}. \]

(4.32)
Also, a.e. on $\Lambda^{N-1}$:
\[
\rho_n^{(N-1)} = \frac{\int_{\Lambda} Q_n(\cdot, x_N) dx_N}{\|Q_n\|_{1,d^N}} \leq \frac{\rho_{n}^{(N-1)} + \frac{1}{n}|\Lambda|}{\|Q_n\|_{1,d^N}} \\
\leq \rho^{(N-1)} + \frac{1}{n}|\Lambda| \leq c + \frac{1}{n}|\Lambda|.
\]
(4.33)

From (4.32) and (4.33) it follows that for every $x_N \in \Lambda$
\[
P_n(\cdot, x_N) \geq \alpha_n \rho_n^{(N-1)}(\cdot) \text{ a.e on } \Lambda^{N-1},
\]
(4.34)
with $\alpha_n := \left[(n + |\Lambda|^N)(c + \frac{1}{n}|\Lambda|)\right]^{-1}$. Thus, (4.14) is satisfied by $P_n$ with $A = \Lambda$, and $\gamma \equiv \alpha_n$. □

Acknowledgements

The author is deeply indebted to Patrick J. Rabier for his generous contributions to this work. The author is glad to express her gratitude to William Noid for first introducing her to the inverse problem.

References

[1] Chayes, J. T., Chayes, L., and Lieb, E. H.: The inverse problem in classical statistical mechanics. *Comm. Math. Physics* **93**, 57-121 (1984)
[2] Evans, R.: The nature of the liquid-vapour interface and other topics in the statistical mechanics of non-uniform, classical fluids. *Adv. Phys.* **28**, 143-200 (1979)
[3] Folland, G.: *Real analysis*. Wiley, New York (1999)
[4] Hoeffding, W.: A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.* **19** 293-325 (1948)
[5] Koroljuk, V. S., Borovskich, Yu. V.: *Theory of U-statistics*. Kluwer Academic Publishers, Dordrecht (1994)
[6] Lee, A. J.: *U-statistics: theory and practice*. Marcel Dekker, New York (1990)
[7] Lyubartsev, A. P. and Laaksonen, A.: Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **52**, 3730-3737 (1995)
[8] Navrotskaya, I.: The classical inverse problem for multi-particle densities in the canonical ensemble formulation. *Preprint* (2015), http://arxiv.org/abs/1406.7345
[9] Noid, W. G.: Systematic methods for structurally consistent coarse-grained models. *Methods. Mol. Biol.* **924**, 487-531 (2013)