Not All Expectations Are Created Linear: Inequalities for Independent Vectors under Sublinear Expectation

Nahom Seyoum

2024-11-07

Contents

Introduction

Classical inequalities involving sample means, such as those related to the law of large numbers or central limit theorems, are well-studied under linear expectations. However, in settings involving uncertainty and model ambiguity, the assumption of linearity is too restrictive. This has motivated a growing interest in sublinear expectations, which allow for a broader class of measures and can better capture uncertainty.

A key development in this area is the representation of a sublinear expectation $\hat{E}$ as the supremum over a convex, weakly compact family of probability measures $\mathcal{P}$. In this framework, the distribution of a random vector X under $\hat{E}$ corresponds to a range of possible linear expectations $\{E_P[X] : P \in \mathcal{P}\}$, forming a convex and compact subset of $\mathbb{R}^d$. By applying the separation theorem, each sublinear expectation induces a unique convex, compact set $\Theta_i \subset \mathbb{R}^d$ such that

$$\Theta_i = \{E_P[X_i] : P \in \mathcal{P}\}.$$

In this paper[5], we consider independent $\mathbb{R}^d$-valued random vectors $\{X_i\}_{i=1}^n$ under a regular sublinear expectation. Our first main contribution is a straightforward argument—avoiding any polytope assumptions—to show that the set Θ_i fully characterizes the expectations of each X_i. Building on this, we establish inequalities involving the sample mean $\frac{1}{n}\sum_{i=1}^n X_i$. In particular, defining

$$\Theta = \left\{ \frac{1}{n} \sum_{i=1}^{n} \theta_i : \theta_i \in \Theta_i \right\}$$

and a suitable “distance”

$$\rho_\Theta(x) = \inf_{\theta \in \Theta} |x - \theta|,$$

we obtain an inequality of the form

$$\hat{E}\left[\rho_\Theta^2 \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right) \right] \leq \frac{\overline{\sigma}_n^2}{n},$$

where

$$\overline{\sigma}_n^2 = \sup_{i \leq n} \inf_{\theta \in \Theta_i} \hat{E}[|X_i - \theta|^2].$$

Further application of Sion’s minimax theorem and Lusin’s theorem yields a minimax characterization of the infimal variance bound over random vectors ξ taking values in Θ.

Preliminary Results

Definition of Sublinear Expectation

Below are the key properties defining a regular sublinear expectation $\widehat{E} : H \to \mathbb{R}$ on $H = \text{Cb.Lip}(\Omega)$.

1. Monotonicity: If $X(\omega) \geq Y(\omega)$ for all ω, then $\widehat{E}[X] \geq \widehat{E}[Y]$.
   If one function is always greater than another, its expected value should be no smaller.
2. Constant Preserving: For any constant c, $\widehat{E}[c] = c$.
   A sure payoff equals its own expectation.

3. Subadditivity: For all X, Y, $\widehat{E}[X+Y] \leq \widehat{E}[X] + \widehat{E}[Y]$.
   The expectation is "conservative": combining risks does not reduce the total risk below the sum of individual parts.

4. Positive Homogeneity: For $\lambda \geq 0$, $\widehat{E}[\lambda X] = \lambda \widehat{E}[X]$.
   Scaling a variable by a nonnegative factor scales its expectation by the same amount.

5. Regularity: If $X_n \downarrow 0$ pointwise, then $\widehat{E}[X_n] \downarrow 0$.
   As a sequence of random variables decreases to zero, their expectations also decrease to zero, ensuring a form of continuity.

The first preliminary theorem we establish is the following:

Theorem 2.1

There exists a convex and weakly compact set of probability measures $\mathcal{P}$ on $(\Omega, \mathcal{B}(\Omega))$ such that:

$$\hat{\mathbb{E}}[X] = \sup_{P \in \mathcal{P}} \mathbb{E}_P[X], \quad \text{for } X \in \mathcal{H},$$

where $\mathcal{B}(\Omega)$ is the Borel σ-field.

They do not present a proof for it but claim it follows naturally from sources 1 and 4. It was initially not intuitive but this is how we reconstructed it.

Proof(2.1)

We begin with $\hat{\mathbb{E}} : \mathcal{H} \to \mathbb{R}$ having the sublinear properties outlined above. Our goal is to write $\widehat{E}[X]$ as $\sup_L L[X]$, where each L is linear and satisfies $L[X] \leq \widehat{E}[X]$ for all $X \in H$. To achieve this, we use a version of the Hahn–Banach theorem adapted for sublinear dominance:

Let $p : V \to \mathbb{R}$ be a sublinear functional on a real vector space V. Suppose $W \subset V$ is a linear subspace and $f : W \to \mathbb{R}$ is a linear functional with $f(x) \leq p(x)$ for all $x \in W$. Then there exists a linear extension $F : V \to \mathbb{R}$ of f such that $F(x) \leq p(x)$ for all $x \in V$.

Apply this lemma with $p = \widehat{E}$ and start from simple linear functionals defined on small subspaces (e.g., just constants). By iterating Zorn's lemma and extending step by step, we construct a family L of linear functionals $L : H \to \mathbb{R}$ such that each L is positive (due to monotonicity preservation) and dominated by $\widehat{E}$:

$$L[X] \leq \widehat{E}[X], \quad \forall X \in H.$$

Moreover, this construction guarantees

$$\widehat{E}[X] = \sup_{L \in L} L[X], \quad \forall X \in H.$$

Each $L \in L$ is linear, positive, and satisfies a continuity-from-above condition inherited from the regularity of $\widehat{E}$. By the Daniell–Stone representation theorem, there exists a unique probability measure P_L on $(\Omega, \mathcal{B}(\Omega))$ such that

$$L[X] = E_{P_L}[X] = \int_\Omega X(\omega) P_L(d\omega), \quad \forall X \in H.$$

Hence each $L \in L$ corresponds to a probability measure P_L. Set $\mathcal{P} := \{P_L : L \in L\}$. Then we have

$$\widehat{E}[X] = \sup_{P \in \mathcal{P}} E_P[X].$$

If $P_1, P_2 \in \mathcal{P}$, then for any $\alpha \in [0,1]$, $\alpha P_1 + (1-\alpha)P_2$ also induces a linear functional $X \mapsto \alpha E_{P_1}[X] + (1-\alpha)E_{P_2}[X]$ which is still $\leq \widehat{E}[X]$. Thus $\alpha P_1 + (1-\alpha)P_2 \in \mathcal{P}$, showing $\mathcal{P}$ is convex. To show $\mathcal{P}$ is weakly compact, we first prove tightness. By the structure of $\widehat{E}$, for any $\epsilon > 0$, we can find a compact set $K_\epsilon \subset \Omega$ such that

$$P(K_\epsilon) \geq 1 - \epsilon, \quad \forall P \in \mathcal{P}.$$

This is because if mass escaped arbitrarily, it would contradict the regularity and monotonicity conditions on $\widehat{E}$. Thus $\mathcal{P}$ is a tight family of probability measures. By Prokhorov's theorem, a tight family of probability measures on a Polish space (and Ω, being complete and separable, is Polish) is relatively weakly compact. Hence every sequence in $\mathcal{P}$ has a weakly convergent subsequence. Let $(P_n)_{n\geq 1}$ be a sequence in $\mathcal{P}$ that converges weakly to some probability measure P_*. For each $X \in H$, since $E_{P_n}[X] \leq \widehat{E}[X]$ and H consists of bounded continuous functions, weak convergence yields $E_{P_n}[X] \to E_{P_*}[X]$. By taking limits and using that $\widehat{E}[X]$ is the supremum over all $\mathcal{P}$-expectations, we get:

$$E_{P_*}[X] \leq \widehat{E}[X], \quad \forall X \in H.$$

This shows P_* also induces a linear functional dominated by $\widehat{E}$. By the construction of $\mathcal{P}$, such a limit measure must also lie in $\mathcal{P}$. Thus, $\mathcal{P}$ is not only relatively weakly compact but also closed under weak limits, ensuring $\mathcal{P}$ is weakly compact.

Now we need to set up other important properties of this space and the expectations.

Construction of $L_p(\Omega)$-Spaces

For each fixed $p \geq 1$, define a norm on $\mathcal{H}$ by

$$\|X\|_p := \left(\hat{\mathbb{E}}\left[|X|^p\right]\right)^{1/p}, \quad X \in \mathcal{H}.$$

Since $\hat{\mathbb{E}}$ is monotone and positively homogeneous, and since $\hat{\mathbb{E}}[|X|^p]$ is finite for all $X \in \mathcal{H}$ (bounded and Lipschitz implies boundedness of X, ensuring finiteness of $\hat{\mathbb{E}}[|X|^p]$ for every p), $\|\cdot\|_p$ is a well-defined norm on $\mathcal{H}$.

By taking the completion of $\mathcal{H}$ with respect to the $\|\cdot\|_p$-norm, we obtain a Banach space $L_p(\Omega)$. Elements of $L_p(\Omega)$ are equivalence classes of Cauchy sequences in $\mathcal{H}$ under the $\|\cdot\|_p$-norm. Formally,

$$L_p(\Omega) = \overline{\mathcal{H}}^{\|\cdot\|_p}.$$

Hölder's inequality in this sublinear context ensures that if $p \geq 1$, then $L_p(\Omega)$ continuously embeds into $L_1(\Omega)$. In other words, for all $p \geq 1$,

$$L_p(\Omega) \subseteq L_1(\Omega).$$

This embedding is justified by an inequality of the form:

$$\hat{\mathbb{E}}\left[|XY|\right] \leq \left(\hat{\mathbb{E}}[|X|^p]\right)^{1/p}\left(\hat{\mathbb{E}}[|Y|^q]\right)^{1/q},$$

with $\frac{1}{p} + \frac{1}{q} = 1$, showing that any p-integrable random variable is also 1-integrable.

Extension of $\hat{\mathbb{E}}$ to $L_1(\Omega)$

Since $\mathcal{H}$ is dense in $L_1(\Omega)$ and $\hat{\mathbb{E}}$ is initially defined on $\mathcal{H}$, we can extend $\hat{\mathbb{E}}$ to all of $L_1(\Omega)$. By the completion process, every $X \in L_1(\Omega)$ is the limit of a sequence $(X_n)_{n=1}^{\infty} \subset \mathcal{H}$ such that $\|X_n - X\|_1 \to 0$. Define:

$$\hat{\mathbb{E}}[X] := \lim_{n\to\infty} \hat{\mathbb{E}}[X_n].$$

This limit is well-defined and does not depend on the choice of the approximating sequence because if (Y_n) is another sequence with $Y_n \to X$ in $L_1(\Omega)$, then $\|X_n - Y_n\|_1 \to 0$, and hence $\hat{\mathbb{E}}[|X_n - Y_n|] \to 0$. By monotonicity and subadditivity, $\hat{\mathbb{E}}[X_n] - \hat{\mathbb{E}}[Y_n] \to 0$, showing consistency.

In addition, $\hat{\mathbb{E}}$ is still a regular sublinear expectation on $L_1(\Omega)$. Regularity, monotonicity, and the other properties extend to this bigger space since limits are taken in the $\|\cdot\|_1$-norm.

Inclusion of $C_b(\Omega)$ in $L_1(\Omega)$

The space $C_b(\Omega)$ denotes all bounded and continuous functions $\Omega \to \mathbb{R}$. By the Stone–Weierstrass theorem (or related approximation theorems), for any bounded continuous $f : \Omega \to \mathbb{R}$ and for any $\epsilon > 0$, there exists a sequence of functions in $\mathcal{H} = C_b^{\text{Lip}}(\Omega)$ that converge uniformly to f. Uniform convergence plus boundedness ensure convergence in the $\|\cdot\|_1$-norm because:

$$\|f-g\|_1 = \left(\hat{\mathbb{E}}[|f-g|]\right)^{1/1} \leq \hat{\mathbb{E}}[\|f-g\|_\infty] = \|f-g\|_\infty$$

for bounded functions f, g. Thus $f \in L_1(\Omega)$. This shows that:

$$C_b(\Omega) \subseteq L_1(\Omega).$$

Since $\mathcal{H} = C_b^{\text{Lip}}(\Omega) \subseteq C_b(\Omega)$, we also have $\mathcal{H} \subseteq L_1(\Omega)$.

Random Vectors in $L_p(\Omega; \mathbb{R}^d)$

A d-dimensional random vector $X = (X_1, \ldots, X_d)$ is said to be in $L_p(\Omega; \mathbb{R}^d)$ if and only if each $X_i \in L_p(\Omega)$. This means:

$$\hat{\mathbb{E}}[|X_i|^p] < \infty \quad \text{for all } i = 1, \ldots, d.$$

We can define a norm on $L_p(\Omega; \mathbb{R}^d)$ by:

$$\|X\|_p := \left(\sum_{i=1}^{d} \hat{\mathbb{E}}[|X_i|^p]\right)^{1/p}.$$

This makes $L_p(\Omega; \mathbb{R}^d)$ a Banach space and again, since $p \geq 1$, we have:

$$L_p(\Omega; \mathbb{R}^d) \subseteq L_1(\Omega; \mathbb{R}^d).$$

Distribution Functionals Under $\hat{\mathbb{E}}$

Given a random vector $X \in L_1(\Omega; \mathbb{R}^d)$, define its distribution under $\hat{\mathbb{E}}$ as a functional:

$$F_X^{\hat{\mathbb{E}}} : C_b^{\text{Lip}}(\mathbb{R}^d) \to \mathbb{R}, \quad F_X^{\hat{\mathbb{E}}}[\phi] := \hat{\mathbb{E}}[\phi(X)].$$

Here, $\phi : \mathbb{R}^d \to \mathbb{R}$ is bounded and Lipschitz. Such ϕ are chosen to ensure that $\phi(X) \in L_1(\Omega)$.

Indepencence of Random Vectors

Although it is not necessarily a result, the definition of independent random vectors in a sublinear setting incredibly interesting. As such, we will walk through components of the definition.

Definition

Let $\{X_i\}_{i=1}^{\infty} \subset L^1(\Omega;\mathbb{R}^d)$, where each $X_i = (X_i^1,\ldots,X_i^d)$ is a d-dimensional random vector with components $X_i^j \in L^1(\Omega)$ (i.e., $\mathbb{E}[|X_i^j|] < \infty$ for all j).

For infinite sequences, the sequence $\{X_i\}_{i=1}^{\infty}$ is called independent if for all $i \geq 1$,

$$\mathbb{E}[\psi(X_1,\ldots,X_i,X_{i+1})] = \mathbb{E}[\mathbb{E}[\psi(x_1,\ldots,x_i,X_{i+1})](x_1,\ldots,x_i = (X_1,\ldots,X_i))],$$

for every test function $\psi \in C_b.\mathrm{Lip}(\mathbb{R}^{d\cdot(i+1)})$.

Similarly, for a finite sequence $\{X_i\}_{i=1}^{n} \subset L^1(\Omega;\mathbb{R}^d)$, where $n > 1$, the sequence is independent if for all $1 \leq i \leq n-1$,

$$\mathbb{E}[\psi(X_1,\ldots,X_i,X_{i+1})] = \mathbb{E}[\mathbb{E}[\psi(x_1,\ldots,x_i,X_{i+1})](x_1,\ldots,x_i = (X_1,\ldots,X_i))],$$

for every test function $\psi \in C_b.\mathrm{Lip}(\mathbb{R}^{d\cdot(i+1)})$.

Interpretation of the Independence Property

Let $\psi(x_1,\ldots,x_i,x_{i+1})$ represent a function of the joint random vector $(X_1,\ldots,X_i,X_{i+1})$. The independence condition states that the expectation of ψ under $\mathbb{E}$ can be decomposed as:

$$\mathbb{E}[\psi(X_1,\ldots,X_i,X_{i+1})] = \mathbb{E}[\mathbb{E}[\psi(x_1,\ldots,x_i,X_{i+1})](x_1,\ldots,x_i = (X_1,\ldots,X_i))].$$

1. Inner Expectation:
   - $\mathbb{E}[\psi(x_1,\ldots,x_i,X_{i+1})]$: This is the expectation of ψ, treating $X_1,\ldots,X_i$ as fixed parameters. It evaluates the influence of X_{i+1} on ψ, conditional on $X_1,\ldots,X_i$.
2. Outer Expectation:
   - $\mathbb{E}[\cdot]$: This computes the expectation over the randomness of $X_1,\ldots,X_i$, which is akin to averaging over all possible realizations of the preceding random vectors.

In classical probability theory, independence is defined using conditional expectations:

$$\mathbb{E}[\psi(X_1,\ldots,X_i,X_{i+1})] = \mathbb{E}[\mathbb{E}[\psi(X_1,\ldots,X_i,X_{i+1})|X_1,\ldots,X_i]].$$

In the sublinear expectation framework, we replace $\mathbb{E}$ with $\mathbb{E}^\wedge$, which may not be additive. The generalization still preserves the notion of "independence" by making sure that X_{i+1}'s distribution is unaffected by the earlier sequence $(X_1,\ldots,X_i)$.

Perhaps the neatest part of the result, in our estimation at least, is the recursive nature of the definition. $(X_1,\ldots,X_i)$ are independent if X_{i+1} is independent of $(X_1,\ldots,X_i)$. This means that the sequence can be studied incrementally (LLN and CLT)!

The final preliminary result we establish is the following proposition:

Proposition 2.3

Let $\{X_i\}_{i=1}^{n}$ be independent random vectors in $L^2(\Omega;\mathbb{R}^d)$ under $\widehat{\mathbb{E}}$. Then, for each $P \in \mathcal{P}$ and $\phi \in \mathrm{C}_{\mathrm{Lip}}(\mathbb{R}^d)$, we have

$$\mathbb{E}^P[\phi(X_i) \mid \mathcal{F}_{i-1}] \leq \widehat{\mathbb{E}}[\phi(X_i)], \quad P\text{-a.s., for } i \leq n,$$

where P is given in Theorem 2.1, $\mathrm{C}_{\mathrm{Lip}}(\mathbb{R}^d)$ denotes the space of Lipschitz functions on $\mathbb{R}^d$, $\mathcal{F}_i = \sigma(X_1,\ldots,X_i)$ for $i \geq 1$, and $\mathcal{F}_0 = \{\emptyset,\Omega\}$.

Again, there is no proof presented for this in the paper apart from a reference to a paper but we have reconstructed it below using some machinary from source 6.

Proof (2.3)

Define the event:

$$B := \left\{ E_P\left[\phi(X_i) \mid F_{i-1}\right] > \hat{E}\left[\phi(X_i)\right] \right\}.$$

Since $E_P\left[\phi(X_i) \mid F_{i-1}\right]$ is F_{i-1}-measurable, we know $B \in F_{i-1}$. We aim to show $P(B) = 0$ for each $P \in \mathcal{P}$.

Suppose, for the sake of contradiction, that there exists some $P \in \mathcal{P}$ with $P(B) > 0$.

Because $B \in F_{i-1}$, there exists a Borel set $F \subseteq \mathbb{R}^{d(i-1)}$ such that:

$$B \supseteq \{(X_1, \ldots, X_{i-1}) \in F\},$$

and hence:

$$P\left(X_1, \ldots, X_{i-1} \in F\right) \geq P(B) > 0.$$

On the event $\{X_1, \ldots, X_{i-1} \in F\}$, we have:

$$E_P\left[\phi(X_i) \mid X_1, \ldots, X_{i-1}\right] > \hat{E}\left[\phi(X_i)\right].$$

The set F is measurable, but the indicator function $\mathbb{I}_F(x_1, \ldots, x_{i-1})$ is generally not Lipschitz. Since our functions need to be in $C^{\text{Lip}}(\mathbb{R}^d)$ or at least constructed from such, we approximate $\mathbb{I}_F$ using bounded Lipschitz functions. This is a standard functional analytic argument:

By Urysohn's lemma and standard approximation arguments, or by Tietze's extension theorem, we can construct a sequence of functions $(\phi_k)_{k\geq 1} \subset C^{\text{Lip}}(\mathbb{R}^{d(i-1)})$ such that:

$$0 \leq \phi_k(x_1, \ldots, x_{i-1}) \leq 1 \quad \text{for all } k,$$

and:

$$\phi_k(x_1, \ldots, x_{i-1}) \downarrow \mathbb{I}_F(x_1, \ldots, x_{i-1}) \quad \text{pointwise as } k \to \infty.$$

This approximation ensures that we can uniformly approximate the indicator of F by bounded Lipschitz functions from above.

We know that on F, the conditional expectation under P of $\phi(X_i)$ is strictly greater than $\hat{E}[\phi(X_i)]$. To exploit this, we want to link this inequality to a contradiction involving the definition of $\hat{E}$.

Let us define a suitable "centering" that will help create a contradiction. Since ϕ is bounded Lipschitz, $\hat{E}[\phi(X_i)]$ is well-defined and finite. Let:

$$m := \hat{E}[\phi(X_i)].$$

Consider the shifted random variable $\phi(X_i) - m$. Its $\hat{E}$-expectation is at most zero by definition (since $\hat{E}$ is sublinear and $\hat{E}[\phi(X_i)] = m$, we have $\hat{E}[\phi(X_i)-m] \leq m-m = 0$). Actually, since $\hat{E}[\phi(X_i)] = m$, there exists some $P^* \in \mathcal{P}$ for which $E_{P^*}[\phi(X_i)]$ approximates m from below or equals it—this ensures $\hat{E}[\phi(X_i)-m] \leq 0$.

Now define, for each integer $N \geq 1$, a truncated function:

$$f_N(x_i) := ((\phi(x_i) - m) \wedge N) \vee (-N).$$

This ensures $f_N : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz and bounded by N. As $N \to \infty$, $f_N(x_i) \to \phi(x_i) - m$ pointwise. Also, since $\phi(x_i)$ is bounded, choose N large enough so that $f_N(x_i) = \phi(x_i) - m$ for all x_i (no truncation needed if N exceeds the essential bound of ϕ). Thus for sufficiently large N:

$$f_N(X_i) = \phi(X_i) - m.$$

We now define the key test functions:

$$\psi_{k,N}(x_1, \ldots, x_{i-1}, x_i) := \phi_k(x_1, \ldots, x_{i-1}) \cdot f_N(x_i).$$

Each $\psi_{k,N}$ is the product of two bounded Lipschitz functions, hence $\psi_{k,N} \in C^{\text{Lip}}(\mathbb{R}^{di})$.

By definition:

$$\mathbb{E}_P[\psi_{k,N}(X_1, \ldots, X_i)] = \mathbb{E}_P[\phi_k(X_1, \ldots, X_{i-1}) \cdot f_N(X_i)].$$

On the event $\{X_1, \ldots, X_{i-1} \in F\}$, we have

$$\mathbb{E}_P[\phi(X_i) \mid X_1, \ldots, X_{i-1}] > m = \hat{\mathbb{E}}[\phi(X_i)],$$

and hence

$$\mathbb{E}_P[f_N(X_i) \mid X_1, \ldots, X_{i-1}] = \mathbb{E}_P[\phi(X_i) - m \mid X_1, \ldots, X_{i-1}] > 0,$$

(for large enough N, $f_N(X_i) = \phi(X_i) - m$ exactly).

Since $\phi_k \downarrow I_F$, by the monotone convergence theorem (applied under P), we have

$$\lim_{k\to\infty} \mathbb{E}_P[\phi_k(X_1, \ldots, X_{i-1}) f_N(X_i)] = \mathbb{E}_P[I_F(X_1, \ldots, X_{i-1}) f_N(X_i)].$$

Moreover, on $\{X_1, \ldots, X_{i-1} \in F\}$, $I_F(X_1, \ldots, X_{i-1}) = 1$, so

$$\mathbb{E}_P[I_F(X_1, \ldots, X_{i-1}) f_N(X_i)] = \mathbb{E}_P(\mathbb{E}_P[I_F(X_1, \ldots, X_{i-1}) f_N(X_i) \mid X_1, \ldots, X_{i-1}]).$$

Since I_F is $\mathcal{F}_{i-1}$-measurable,

$$\mathbb{E}_P[I_F(X_1, \ldots, X_{i-1}) f_N(X_i) \mid X_1, \ldots, X_{i-1}] = I_F(X_1, \ldots, X_{i-1}) \mathbb{E}_P[f_N(X_i) \mid X_1, \ldots, X_{i-1}].$$

On F, this is strictly positive. Thus

$$\mathbb{E}_P[I_F(X_1, \ldots, X_{i-1}) f_N(X_i)] > 0,$$

and hence for sufficiently large k,

$$\mathbb{E}_P[\psi_{k,N}(X_1, \ldots, X_i)] > 0.$$

Now, recall that $\hat{\mathbb{E}}$ is sublinear and

$$\hat{\mathbb{E}}[X] = \sup_{Q \in \mathcal{P}} \mathbb{E}_Q[X].$$

In particular, for any X, $\mathbb{E}_P[X] \leq \hat{\mathbb{E}}[X]$.

Apply this to $X = \psi_{k,N}(X_1, \ldots, X_i)$:

$$\mathbb{E}_P[\psi_{k,N}(X_1, \ldots, X_i)] \leq \hat{\mathbb{E}}[\psi_{k,N}(X_1, \ldots, X_i)].$$

We also know from the independence property under $\hat{\mathbb{E}}$ that since $\psi_{k,N}$ factors into a function of $(X_1, \ldots, X_{i-1})$ times a function of X_i, we have

$$\hat{\mathbb{E}}[\psi_{k,N}(X_1, \ldots, X_i)] = \hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1}) f_N(X_i)].$$

By the definition of independence (and considering that $f_N(X_i) = \phi(X_i) - m$), it can be shown that $\hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1}) f_N(X_i)]$ factors into $\hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1})]\hat{\mathbb{E}}[f_N(X_i)]$ or is at most $\hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1})] \max(\hat{\mathbb{E}}[f_N(X_i)], 0)$ due to sublinearity and the "independence" structure.

Since ϕ_k is bounded and approximates an indicator, $\hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1})]$ is finite and converges to $\hat{\mathbb{E}}[I_F(X_1, \ldots, X_{i-1})]$.

Crucially, because $\hat{\mathbb{E}}[\phi(X_i)] = m$, we have $\hat{\mathbb{E}}[\phi(X_i) - m] \leq 0$. By the boundedness of ϕ, as N grows large enough so that no truncation occurs, $\hat{\mathbb{E}}[f_N(X_i)] = \hat{\mathbb{E}}[\phi(X_i) - m] \leq 0$.

Thus,

$$\hat{\mathbb{E}}[\psi_{k,N}(X_1, \ldots, X_i)] \leq \hat{\mathbb{E}}[\phi_k(X_1, \ldots, X_{i-1})] \cdot 0 = 0.$$

This implies

$$\mathbb{E}_P[\psi_{k,N}(X_1, \ldots, X_i)] \leq 0.$$

But previously, we derived that for large enough k, N,

$$\mathbb{E}_P[\psi_{k,N}(X_1, \ldots, X_i)] > 0.$$

This is a direct contradiction. Since $P \in \mathcal{P}$ was arbitrary, we conclude that for every $P \in \mathcal{P}$,

$$\mathbb{E}_P[\phi(X_i) \mid \mathcal{F}_{i-1}] \leq \hat{\mathbb{E}}[\phi(X_i)], \quad P\text{-a.s.}$$

Main Results

There are three major results established in this paper. Theorem 3.1 identifies the set Θ_i as exactly the collection of all possible expectations $\mathbb{E}_P[X_i]$ for $P \in \mathcal{P}$, and shows that the conditional expectations $\mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}]$ also lie in Θ_i almost surely. Theorem 3.2 bounds the "variance" of the sample mean under sublinear expectations by $\frac{\overline{\sigma}_n^2}{n}$ where $\overline{\sigma}_n^2$ is a measure of dispersion controlled by the sets Θ_i. Finally, theorem 3.4 , using Sion's minimax theorem and $L^2(\Omega; \Theta)$-approximations, establishes that the order of taking an infimum over ξ and a supremum over P can be interchanged, which gives us a minimax characterization of the best achievable variance bound within Θ.

Before outlining their proofs in more detail, we will first establish some machinary to assist us.

Let $\{X_i\}_{i=1}^n$ be independent random vectors in $L^2(\Omega; \mathbb{R}^d)$ under $\hat{\mathbb{E}}$. For each $i \leq n$, define $g_i : \mathbb{R}^d \to \mathbb{R}$ as follows:

Let

$$g_i(p) := \mathbb{E}[\langle p, X_i \rangle], \quad p \in \mathbb{R}^d.$$

For $p_1, p_2 \in \mathbb{R}^d$:

$$g_i(p_1 + p_2) = \mathbb{E}[\langle p_1 + p_2, X_i \rangle] = \mathbb{E}[\langle p_1, X_i \rangle + \langle p_2, X_i \rangle].$$

By the subadditivity of $\mathbb{E}$:

$$g_i(p_1 + p_2) \leq \mathbb{E}[\langle p_1, X_i \rangle] + \mathbb{E}[\langle p_2, X_i \rangle] = g_i(p_1) + g_i(p_2).$$

For $\lambda \geq 0$ and $p \in \mathbb{R}^d$:

$$g_i(\lambda p) = \mathbb{E}[\langle \lambda p, X_i \rangle] = \mathbb{E}[\lambda \langle p, X_i \rangle].$$

By positive homogeneity of $\mathbb{E}$:

$$g_i(\lambda p) = \lambda \mathbb{E}[\langle p, X_i \rangle] = \lambda g_i(p).$$

Thus, g_i is subadditive and positively homogeneous.

Theorem 1.2.1 from source 6 states that any sublinear expectation can be represented as the supremum of linear expectations. Applying this to the linear functionals $p \mapsto \langle p, X_i \rangle$, we conclude:

$$g_i(p) = \mathbb{E}[\langle p, X_i \rangle] = \sup_{\theta \in \Theta_i} \langle p, \theta \rangle.$$

From the representation:

$$g_i(p) = \sup_{\theta \in \Theta_i} \langle p, \theta \rangle,$$

we deduce that for each fixed θ:

$$\langle \theta, p \rangle \leq g_i(p) \quad \text{for all } p \in \mathbb{R}^d.$$

Thus,

$$\Theta_i = \{\theta \in \mathbb{R}^d : \langle \theta, p \rangle \leq g_i(p) \text{ for all } p \in \mathbb{R}^d\}.$$

We can now check for the convexity, compactness and uniqueness properties.

Θ_i is defined by linear inequalities of the form $\langle \theta, p \rangle \leq g_i(p)$. Each inequality defines a half-space, and the intersection of half-spaces is convex. Therefore, Θ_i is convex.

Since g_i is sublinear and finite for all p, it dominates $\langle p, \theta \rangle$. This bounds the coordinates of θ, ensuring Θ_i is bounded. Being closed (as an intersection of closed half-spaces), convex, and bounded in finite-dimensional space, Θ_i is compact.

The set Θ_i is uniquely determined by g_i because any other set producing the same support function g_i must coincide with Θ_i. This follows from the uniqueness of support functions for convex bodies.

Theorem 3.1

Let $\{X_1, X_2, \ldots, X_n\}$ be independent random vectors in $L^2(\Omega; \mathbb{R}^d)$ under $\hat{E}$. Then we have:

1. $\Theta_i = \{\mathbb{E}_P[X_i] : P \in \mathcal{P}\}$ for $i \leq n$, where $\mathcal{P}$ is given in Theorem 2.1, and Θ_i is defined in (3.3).
2. For each $P \in \mathcal{P}$ and $i \leq n$, $\mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}] \in \Theta_i$, P-a.s., where $\mathcal{F}_i = \sigma(X_1, \ldots, X_i)$ for $i \geq 1$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

Proof (3.1.1)

Define

$$\tilde{\Theta}_i := \{\mathbb{E}_P[X_i] : P \in \mathcal{P}\}.$$

Since $\mathcal{P}$ is convex, for any $P_1, P_2 \in \mathcal{P}$ and $\lambda \in [0,1]$:

$$\lambda \mathbb{E}_{P_1}[X_i] + (1-\lambda)\mathbb{E}_{P_2}[X_i] = \mathbb{E}_{\lambda P_1 + (1-\lambda)P_2}[X_i],$$

where $\lambda P_1 + (1-\lambda)P_2 \in \mathcal{P}$ by convexity. Thus, $\tilde{\Theta}_i$ is convex.

By assumption, $\mathcal{P}$ is weakly compact. The map $P \mapsto \mathbb{E}_P[X_i]$ is continuous with respect to the weak convergence of probability measures. Hence, $\tilde{\Theta}_i$ is the continuous image of a compact set $\mathcal{P}$, thus $\tilde{\Theta}_i$ is compact.

Therefore, $\tilde{\Theta}_i$ is a nonempty, convex, and compact subset of $\mathbb{R}^d$.

Take any $y \in \tilde{\Theta}_i$. By definition, there exists $P_y \in \mathcal{P}$ such that $y = \mathbb{E}_{P_y}[X_i]$.

For all $p \in \mathbb{R}^d$:

$$\langle p, y \rangle = \langle p, \mathbb{E}_{P_y}[X_i] \rangle = \mathbb{E}_{P_y}[\langle p, X_i \rangle] \leq \sup_{P \in \mathcal{P}} \mathbb{E}_P[\langle p, X_i \rangle] = g_i(p).$$

Since this holds for all p, we have $y \in \Theta_i$. Thus, $\tilde{\Theta}_i \subseteq \Theta_i$.

By definition:

$$g_i(p) = \sup_{P \in \mathcal{P}} \mathbb{E}_P[\langle p, X_i \rangle].$$

On the other hand, from the set $\tilde{\Theta}_i$:

$$\sup_{y \in \tilde{\Theta}_i} \langle p, y \rangle = \sup_{P \in \mathcal{P}} \langle p, \mathbb{E}_P[X_i] \rangle = \sup_{P \in \mathcal{P}} \mathbb{E}_P[\langle p, X_i \rangle] = g_i(p).$$

Thus, g_i is precisely the support function of the compact convex set $\tilde{\Theta}_i$:

$$g_i(p) = h_{\tilde{\Theta}_i}(p) := \sup_{y \in \tilde{\Theta}_i} \langle p, y \rangle.$$

For a compact, convex subset $C \subset \mathbb{R}^d$, the support function h_C uniquely determines C. In particular, if another set D satisfies $h_D = h_C$, then $C = D$.

Since Θ_i is defined by:

$$\Theta_i = \{y : \langle p, y \rangle \leq g_i(p) \ \forall p \in \mathbb{R}^d\},$$

it is the maximal closed, convex set whose support function is g_i.

We have identified one such set with support function g_i, namely $\tilde{\Theta}_i$. By uniqueness, $\Theta_i = \tilde{\Theta}_i$.

Proof (3.1.2)

Let $P \in P$ and fix $1 \leq i \leq n$.

Define $Q := Q$ and

$$Q_d := \{(p_1, \ldots, p_d) \in \mathbb{R}^d : p_j \in Q \text{ for each } j = 1, \ldots, d\}.$$

Note that Q_d is countable and dense in $\mathbb{R}^d$.

From Proposition 2.3, for each $p \in Q_d$,

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}] \rangle = \mathbb{E}_P[\langle p, X_i \rangle \mid \mathcal{F}_{i-1}] \leq g_i(p) \quad P\text{-a.s.}$$

This means: for each $p \in Q_d$, there exists a P-null set $N_p \subset \Omega$ such that

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega) \rangle \leq g_i(p) \quad \forall \omega \in \Omega \setminus N_p.$$

We can begin by constructing a universal null set for all rational directions

We have a family of null sets $\{N_p : p \in Q_d\}$. Since Q_d is countable, write $Q_d = \{p_1, p_2, p_3, \ldots\}$.

Consider

$$N := \bigcup_{k=1}^{\infty} N_{p_k}.$$

Since each N_{p_k} is a P-null set, and a countable union of null sets has measure zero, we have:

$$P(N) = 0.$$

On $\Omega \setminus N$, for all $p \in Q_d$ simultaneously,

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \le g_i(p).$$

Thus, we have enforced that one single null set N works for every rational vector p.

Now we extend the inequality in all real directions:

Let $\omega \in \Omega \setminus N$. We know:

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \le g_i(p) \quad \text{for all } p \in Q_d.$$

Take arbitrary $p \in \mathbb{R}^d$. Since Q_d is dense in $\mathbb{R}^d$, there exists a sequence $(p^{(m)})_{m=1}^{\infty} \subset Q_d$ such that $p^{(m)} \to p$ as $m \to \infty$.

For each m:

$$\langle p^{(m)}, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \le g_i(p^{(m)}).$$

As $m \to \infty$, the left side

$$\langle p^{(m)}, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \to \langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle$$

by continuity of the inner product and the pointwise convergence $p^{(m)} \to p$.

To pass the limit on the right side:

$$g_i(p) = \sup_{P' \in P} \mathbb{E}_{P'}[\langle p, X_i\rangle].$$

The function g_i is convex (as a supremum of linear forms) and hence continuous from below on $\mathbb{R}^d$. In particular, since $p^{(m)} \to p$, we have:

$$\lim_{m\to\infty} g_i(p^{(m)}) = g_i(p).$$

Combining these:

$$\lim_{m\to\infty} \langle p^{(m)}, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle = \langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle,$$

and

$$\lim_{m\to\infty} g_i(p^{(m)}) = g_i(p).$$

By taking limits, we preserve the inequality, hence:

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \le g_i(p).$$

Since $p \in \mathbb{R}^d$ was arbitrary, we have shown:

$$\langle p, \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega)\rangle \le g_i(p) \quad \forall p \in \mathbb{R}^d.$$

This holds for all $\omega \in \Omega \setminus N$, and we recall $P(N) = 0$. By definition:

$$\Theta_i = \{y \in \mathbb{R}^d : \langle p, y \rangle \leq g_i(p) \text{ for all } p \in \mathbb{R}^d\}.$$

For each $\omega \in \Omega \setminus N$, we have established:

$$\mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}](\omega) \in \Theta_i.$$

Since $P(N) = 0$, we have:

$$\mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}] \in \Theta_i \quad P\text{-a.s.}$$

The penultimate result we will prove is the following:

Theorem 3.2

Let $\{X_i\}_{i=1}^n$ be independent random vectors in $L^2(\Omega; \mathbb{R}^d)$ under the sublinear expectation $\hat{E}$. Define

$$\Theta := \left\{ \frac{1}{n} \sum_{i=1}^n \theta_i : \theta_i \in \Theta_i \text{ for each } i = 1, \ldots, n \right\},$$

where each Θ_i is defined as in (3.3). For $x \in \mathbb{R}^d$, set

$$\rho_\Theta(x) := \inf_{\theta \in \Theta} |x - \theta|.$$

Define

$$\overline{\sigma}_n^2 := \sup_{1 \leq i \leq n} \inf_{\theta_i \in \Theta_i} \hat{E}[|X_i - \theta_i|^2].$$

Then:

$$\hat{E}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \leq \frac{\overline{\sigma}_n^2}{n}.$$

Proof (3.2)

By Theorem 2.1, there is a convex, weakly compact set P of probability measures such that for any random variable Y,

$$\hat{E}[Y] = \sup_{P \in \mathcal{P}} E_P[Y].$$

Apply this to $Y = \rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)$:

$$\hat{E}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] = \sup_{P \in \mathcal{P}} E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right].$$

From Theorem 3.1(2), for each fixed $P \in \mathcal{P}$ and each $i \leq n$:

$$E_P[X_i | F_{i-1}] \in \Theta_i \quad P\text{-a.s.}$$

Since $E_P[X_i | F_{i-1}]$ is F_{i-1}-measurable and Θ_i-valued a.s., their average

$$\frac{1}{n}\sum_{i=1}^n E_P[X_i | F_{i-1}]$$

is also Θ-valued P-a.s. Thus there exists a random vector

$$\theta(\omega) := \frac{1}{n}\sum_{i=1}^{n} E_P[X_i|F_{i-1}](\omega)$$

such that $\theta(\omega) \in \Theta$ for almost all ω.

By definition of ρ_Θ, for each $\omega \in \Omega$:

$$\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i(\omega)\right) = \inf_{\theta\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) - \theta\right|^2.$$

Since $\theta(\omega) \in \Theta$ a.s., we have pointwise a.s.:

$$\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i(\omega)\right) \le \left|\frac{1}{n}\sum_{i=1}^{n} X_i(\omega) - \frac{1}{n}\sum_{i=1}^{n} E_P[X_i|F_{i-1}](\omega)\right|^2.$$

Taking expectation under P:

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \le E_P\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E_P[X_i|F_{i-1}])\right|^2\right].$$

For any finite set of vectors $y_1, \ldots, y_n$:

$$\left|\frac{1}{n}\sum_{i=1}^{n} y_i\right|^2 \le \frac{1}{n^2}\sum_{i=1}^{n}|y_i|^2.$$

Apply this with $y_i = X_i - E_P[X_i|F_{i-1}]$:

$$E_P\left[\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E_P[X_i|F_{i-1}])\right|^2\right] \le \frac{1}{n^2}\sum_{i=1}^{n} E_P[|X_i - E_P[X_i|F_{i-1}]|^2].$$

Thus:

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \le \frac{1}{n^2}\sum_{i=1}^{n} E_P[|X_i - E_P[X_i|F_{i-1}]|^2].$$

By the properties of conditional expectation in L^2:

$$E_P[|X_i - E_P[X_i|F_{i-1}]|^2] \le E_P[|X_i - E_P[X_i]|^2].$$

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \le \frac{1}{n^2}\sum_{i=1}^{n} E_P[|X_i - E_P[X_i]|^2].$$

By Theorem 3.1(1), $E_P[X_i] \in \Theta_i$. Hence:

$$E_P[|X_i - E_P[X_i]|^2] = \inf_{\theta_i\in\Theta_i} E_P[|X_i - \theta_i|^2].$$

Thus:

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] \le \frac{1}{n^2}\sum_{i=1}^{n}\inf_{\theta_i\in\Theta_i} E_P[|X_i - \theta_i|^2].$$

Since $\hat{E}[Z] = \sup_{P\in\mathcal{P}} E_P[Z]$ and thus $E_P[Z] \le \hat{E}[Z]$ for all $P \in \mathcal{P}$, we have:

$$\inf_{\theta_i\in\Theta_i} E_P[|X_i - \theta_i|^2] \le \inf_{\theta_i\in\Theta_i} \hat{E}[|X_i - \theta_i|^2].$$

Therefore:

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{1}{n^2}\sum_{i=1}^n \inf_{\theta_i\in\Theta_i} \hat{E}[|X_i - \theta_i|^2].$$

By definition:

$$\overline{\sigma}_n^2 = \sup_{1\le i\le n} \inf_{\theta_i\in\Theta_i} \hat{E}[|X_i - \theta_i|^2].$$

Since it is a supremum, for each i:

$$\inf_{\theta_i\in\Theta_i} \hat{E}[|X_i - \theta_i|^2] \le \overline{\sigma}_n^2.$$

Hence:

$$E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{1}{n^2}\sum_{i=1}^n \overline{\sigma}_n^2 = \frac{\overline{\sigma}_n^2}{n}.$$

Since the inequality holds for all $P \in \mathcal{P}$, taking supremum over P:

$$\sup_{P\in\mathcal{P}} E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{\overline{\sigma}_n^2}{n}.$$

Since the left-hand side supremum is $\hat{E}$:

$$\hat{E}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] = \sup_{P\in\mathcal{P}} E_P\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{\overline{\sigma}_n^2}{n}.$$

Theorem 3.4

Let $\{X_i\}_{i=1}^n$ be independent random vectors in $L^2(\Omega;\mathbb{R}^d)$ under $\hat{E}$.

Set $L^2(\Omega;\Theta) = \{\xi \in L^2(\Omega;\mathbb{R}^d) : \xi(\omega) \in \Theta \text{ for } \omega \in \Omega\}$, where Θ is defined in Theorem 3.2.

Then:

$$\inf_{\xi\in L^2(\Omega;\Theta)} \hat{E}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \xi\right|^2\right] \le \frac{\overline{\sigma}_n^2}{n}.$$

Proof (3.4)

Sion's Minimax Theorem

Let P be the given weakly compact, convex set of probability measures. By definition, for any bounded continuous function f, the map $P \mapsto \mathbb{E}_P[f]$ is continuous under the topology of weak convergence of measures. Since each $X_i \in L^2(\Omega;\mathbb{R}^d)$, the random variables and their squares are integrable, and standard density arguments ensure we can approximate $\left|\frac{1}{n}\sum_i X_i - \xi\right|^2$ by bounded continuous functions in probability, preserving continuity under weak convergence. Define $\Phi : P \times L^2(\Omega;\Theta) \to \mathbb{R}$ by

$$\Phi(P,\xi) := \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - \xi\right|^2\right].$$

Let's look at the properties of Φ

1. **For fixed** ξ, the map $P \mapsto \Phi(P,\xi)$ is linear in P. Since $\mathbb{E}_P$ is linear in P, for $P_\lambda = \lambda P_1 + (1-\lambda)P_2$,

$$\Phi(P_\lambda,\xi) = \mathbb{E}_{P_\lambda}\left[|Z-\xi|^2\right] = \lambda\mathbb{E}_{P_1}[|Z-\xi|^2] + (1-\lambda)\mathbb{E}_{P_2}[|Z-\xi|^2].$$

   Thus, $\Phi(P,\xi)$ is affine in P, and since P is convex, $\sup_{P\in P}\Phi(P,\xi)$ is well-defined.

2. **For fixed** P, the map $\xi \mapsto \Phi(P,\xi)$ is strictly convex. Since

$$\Phi(P,\xi) = \mathbb{E}_P\left[|Z-\xi|^2\right], \quad Z := \frac{1}{n}\sum_{i=1}^{n} X_i,$$

   the L^2-norm ensures strict convexity. For $\xi_1,\xi_2 \in L^2(\Omega;\Theta)$ and $\lambda \in [0,1]$,

$$\Phi(P,\lambda\xi_1 + (1-\lambda)\xi_2) \le \lambda\Phi(P,\xi_1) + (1-\lambda)\Phi(P,\xi_2),$$

   with equality only if $\xi_1 = \xi_2$ a.s.

- Φ is convex in ξ for each fixed P.
- Φ is affine (hence both convex and concave) in P for each fixed ξ.
- P is compact in the weak topology.
- $L^2(\Omega;\Theta)$ is convex.
- Φ is upper semicontinuous in the P-variable under the weak topology and convex in the ξ-variable.

By Sion's minimax theorem, this ensures:

$$\inf_{\xi\in L^2(\Omega;\Theta)}\sup_{P\in P}\Phi(P,\xi) = \sup_{P\in P}\inf_{\xi\in L^2(\Omega;\Theta)}\Phi(P,\xi).$$

Approximation in $L^2(\Omega;\Theta)$

Assume we have a random vector

$$\zeta_P(\omega) := \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_P[X_i|\mathcal{F}_{i-1}](\omega),$$

with $\zeta_P(\omega) \in \Theta$ a.s. and $\zeta_P \in L^2(\Omega;\mathbb{R}^d)$.

We want to approximate ζ_P in $L^2(P)$-norm by a sequence $\{\xi_k\} \subset L^2(\Omega;\Theta)$ with $\xi_k(\omega) \in \Theta$ a.s.

Since Θ is compact in $\mathbb{R}^d$, for any $\delta > 0$, there exists a finite δ-net $\{\theta_j\}_{j=1}^m \subset \Theta$ such that

$$\Theta \subset \bigcup_{j=1}^{m} B(\theta_j,\delta),$$

where $B(\theta_j,\delta)$ denotes the ball of radius δ around θ_j.

For almost every ω, $\zeta_P(\omega) \in \Theta$. Thus, there exists at least one $j(\omega)$ for which $\zeta_P(\omega) \in B(\theta_{j(\omega)},\delta)$. Set

$$\xi_\delta(\omega) := \theta_{j(\omega)}.$$

Then $\xi_\delta(\omega) \in \Theta$ a.s., so $\xi_\delta \in L^2(\Omega;\Theta)$ (it is bounded since Θ is compact, hence $\xi_\delta \in L^2(\Omega;\mathbb{R}^d)$).

For all ω,

$$|\zeta_P(\omega) - \xi_\delta(\omega)| \le \delta.$$

Therefore,

$$\mathbb{E}_P[|\zeta_P - \xi_\delta|^2] \le \delta^2.$$

By choosing $\delta = \varepsilon$ for any $\varepsilon > 0$, we get a sequence $\xi_k := \xi_{1/k}$ that converges to ζ_P in $L^2(P)$-norm:

$$\lim_{k\to\infty} \mathbb{E}_P[|\zeta_P - \xi_{1/k}|^2] = 0.$$

Now, we can prove the theorem.

From above, we have

$$\inf_{\xi\in L^2(\Omega;\Theta)} \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi\right|^2\right] = \inf_{\xi\in L^2(\Omega;\Theta)} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi\right|^2\right].$$

By the minimax equality:

$$\inf_{\xi\in L^2(\Omega;\Theta)} \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi\right|^2\right] = \sup_{P\in\mathcal{P}} \inf_{\xi\in L^2(\Omega;\Theta)} \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi\right|^2\right].$$

Now, we construct the approximation using conditional expectations:

For each fixed P, consider:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}] \in \Theta \quad P\text{-a.s.}$$

By the $L^2(\Omega;\Theta)$-approximation argument, we have a sequence $\{\xi_k\} \subset L^2(\Omega;\Theta)$ such that:

$$\lim_{k\to\infty} \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi_k\right|^2\right] = \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}]\right|^2\right].$$

Thus:

$$\inf_{\xi\in L^2(\Omega;\Theta)} \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \xi\right|^2\right] \le \mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}]\right|^2\right].$$

From Theorem 3.2, we know:

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}]\right|^2\right] \le \frac{\bar{\sigma}_n^2}{n}.$$

Since $\mathbb{E}_P[Z] \le \mathbb{E}[Z]$ for all P and random variables Z, we have:

$$\mathbb{E}_P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_P[X_i \mid \mathcal{F}_{i-1}]\right|^2\right] \le \frac{\bar{\sigma}_n^2}{n}.$$

Therefore:

$$\inf_{\xi \in L^2(\Omega;\Theta)} \mathbb{E}_P \left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \xi \right|^2 \right] \leq \frac{\bar{\sigma}_n^2}{n}.$$

Taking supremum over P:

$$\sup_{P \in \mathcal{P}} \inf_{\xi \in L^2(\Omega;\Theta)} \mathbb{E}_P \left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \xi \right|^2 \right] \leq \frac{\bar{\sigma}_n^2}{n}.$$

By the minimax equality:

$$\inf_{\xi \in L^2(\Omega;\Theta)} \sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \xi \right|^2 \right] = \sup_{P \in \mathcal{P}} \inf_{\xi} \mathbb{E}_P[\ldots] \leq \frac{\bar{\sigma}_n^2}{n}.$$

$$\inf_{\xi \in L^2(\Omega;\Theta)} \mathbb{E} \left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \xi \right|^2 \right] = \inf_{\xi \in L^2(\Omega;\Theta)} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\ldots].$$

Hence:

$$\inf_{\xi \in L^2(\Omega;\Theta)} \mathbb{E} \left[\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \xi \right|^2 \right] \leq \frac{\bar{\sigma}_n^2}{n}.$$

Conclusion & Extension

All in all, we investigated the behavior of independent random vectors under a sublinear expectation framework, where uncertainty is modeled by a convex and weakly compact family of probability measures, denoted as P.

We began by showing that each sublinear expectation $\hat{E}$ can be represented as a supremum over linear expectations E_P for $P \in P$. Using this representation, we identified and characterized the sets

$$\Theta_i = \{E_P[X_i] : P \in P\}.$$

These sets Θ_i, derived from the distributional uncertainty, uniquely determine the possible expectations of each X_i. Once the sets Θ_i were established as convex, compact subsets of $\mathbb{R}^d$, we combined them to define:

$$\Theta := \left\{ \frac{1}{n} \sum_{i=1}^{n} \theta_i : \theta_i \in \Theta_i \right\}.$$

Then, we introduced a distance $\rho_\Theta(x)$ to measure how far a point x lies from Θ, defined as:

$$\rho_\Theta(x) := \inf_{\theta \in \Theta} |x - \theta|.$$

Using conditional expectations under measures in P and the properties of sublinear expectations, we derived the inequality:

$$\hat{E}\left[\rho_\Theta^2\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] \le \frac{\bar{\sigma}_n^2}{n},$$

where:

$$\bar{\sigma}_n^2 = \sup_{1\le i\le n} \inf_{\theta_i\in\Theta_i} \hat{E}\left[|X_i - \theta_i|^2\right].$$

Ultimately, by applying Sion's minimax theorem and an $L^2(\Omega;\Theta)$-approximation argument, we established a minimax equivalence that characterizes the least possible "variance" of the sample mean with respect to approximations inside Θ. This leads to a clean minimax expression for the variance bound.

A possible natural extension is to examine sequences $\{X_i\}$ of independent random vectors under sublinear expectations as $n \to \infty$. One might investigate laws of large numbers and central limit theorems in this setting, exploring the asymptotic behavior of $\frac{1}{n}\sum_{i=1}^n X_i$.

Bibliography

[1] Denis, L., Hu, M., & Peng, S. (2011). Function spaces and capacity related to a sublinear expectation: application to G-Brownian motion paths. *Potential Analysis, 34*, 139–161.

[2] Fang, X., Peng, S., Shao, Q., & Song, Y. (2019). Limit theorems with rate of convergence under sublinear expectations. *Bernoulli, 25*(4A), 2564–2596.

[3] Hu, M., Li, X., & Li, X. (2021). Convergence rate of Peng's law of large numbers under sublinear expectations. *Probability, Uncertainty and Quantitative Risk, 6*, 261–266.

[4] Hu, M., & Peng, S. (2009). On representation theorem of G-expectations and paths of G-Brownian motion. *Acta Mathematicae Applicatae Sinica, English Series, 25*, 539–546.

[5] Li, X., & Hu, M. (2024). *Inequalities for independent random vectors under sublinear expectations.* arXiv preprint, arXiv:2410.11349.

[6] Peng, S. (2019). *Nonlinear Expectations and Stochastic Calculus under Uncertainty.* Springer.

[7] Pollard, D. *A User's Guide to Measure Theoretic Probability*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2001.