The existence of the least favorable noise

Dongzhou Huang*

Abstract

Suppose that a random variable X of interest is observed. This paper concerns "the least favorable noise" $\hat{Y}_\epsilon$, which maximizes the prediction error $E[X - E[X|X + Y]]^2$ (or minimizes the variance of $E[X|X + Y]$) in the class of Y with Y independent of X and $\text{var} Y \leq \epsilon^2$. This problem was first studied by Ernst, Kagan, and Rogers ([4]). In the present manuscript, we show that the least favorable noise $\hat{Y}_\epsilon$ must exist and that its variance must be ϵ^2. The proof of existence relies on a convergence result we develop for variances of conditional expectations. Further, we show that the function $\inf_{\text{var} Y \leq \epsilon^2} \text{var} E[X|X + Y]$ is both strictly decreasing and right continuous in ϵ.

1 Introduction

In 2022, Ernst, Kagan and Rogers ([4]) investigated the problem of "the least favorable noise" for an observed square-integrable random variable X. The authors considered $\hat{Y}_\epsilon$, a square-integrable random variable independent of X, which maximizes the prediction error

$$E[X - E[X|X + Y]]^2 = \text{var} X - \text{var} E[X|X + Y],$$

(or, equivalently, minimizes the variance of $E[X|X + Y]$) in the class of Y with $\text{var} Y \leq \epsilon^2$. The authors proceeded to characterize the least favorable noise and show that Y should be the least favorable noise if the distribution of Y satisfies the conditions of a given characterization.

The present manuscript takes a step back from the Ernst et al. characterization of the least favorable noise and asks the following question: 'does the least favorable noise exist?' In other words, given the distribution of X, does there exist a distribution of Y which achieves the maximum of $E[X - E[X|X + Y]]^2$? Although Ernst et al. show the existence of the least favorable noise given (i) the distribution of X and (ii) given that the distribution of Y satisfies three characterization conditions in [4], the conditions are somewhat complicated and, more importantly, nearly impossible to verify for most distributions of X. Therefore, the question of existence of the least favorable noise has remained open. The key contribution of the present paper is to close this problem by showing the existence of the least favorable noise for any distribution of X. The proof

---

*Department of Statistics, Colorado State University, United States of America. E-mail: dongzhou.huang@colostate.edu

relies on a convergence result for the variance of the conditional expectation of X_n given $X_n + Y_n$, provided (X_n, Y_n) converges weakly (see Theorem 2.2).

Before proceeding with the proof of the existence of the least favorable noise, we pause to provide some practical implications of the least favorable noise. In some applied scientific scenarios, the observed random signal X may be highly volatile, making it difficult to analyze. In this case, one may wish to simplify the signal X while keeping its main structure. One possible way to do so is to consider the random variable

$$Q = E\left[X|X + \hat{Y}_\epsilon\right],$$

where $\hat{Y}_\epsilon$ is the least favorable noise. The random variable Q has two important properties: (i) The variance of Q is no more than the variance of X and achieves the minimum of $\operatorname{var} E[X|X + Y]$ over Y satisfying $\operatorname{var} Y \le \epsilon^2$, meaning that it is less volatile, and (ii) noting that $X + \hat{Y}_\epsilon$ is close to X for small ϵ, the random variable Q, as a function of $X + \hat{Y}_\epsilon$, preserves the structure of X.

We now formalize the problem under consideration. To avoid trivialities, we assume X is a non-degenerate and that it is a square-integrable random variable. Without loss of generality, we may assume the mean of X to be 0. Since the conditional expectation $E[X|X + Y]$ remains the same if we shift Y by a constant, we shall only focus on the following class of random variables Y:

$$\mathcal{V}_\epsilon(X) := \{Y : Y \text{ independent of } X,\ E[Y] = 0 \text{ and } E[Y^2] \le \epsilon^2\}.$$

We define

$$L(X, \epsilon) := \inf_{Y \in \mathcal{V}_\epsilon(X)} \operatorname{var} E[X|X + Y].$$

Then the least favorable noise $\hat{Y}_\epsilon$ is a random variable Y in $\mathcal{V}_\epsilon(X)$ such that $\operatorname{var} E[X|X+Y]$ achieves $L(X, \epsilon)$.

In Section 2, we prove the existence of the least favorable noise. That is, we prove the existence of $\hat{Y}_\epsilon$ which minimizes $\operatorname{var} E[X|X + Y]$ when $\operatorname{var} Y \le \epsilon^2$. In general, to show the existence of the minimizer of a value function $f(x)$, one typically adopts the following strategy:

(i) One constructs a sequence $\{x_n\}_{n=1}^\infty$ such that $f(x_n)$ converges to $\inf_x f(x)$;

(ii) One finds a convergent subsequence of $\{x_n\}_{n=1}^\infty$ which converges to x_*;

(iii) One then shows that $f(x_*) = \inf_x f(x)$, from which one concludes that x_* is the minimizer.

In this paper, we indeed follow the above strategy. Firstly, we consider a sequence $\{Y_n\}_{n=1}^\infty$ such that $\operatorname{var} Y_n \le \epsilon^2$ and such that the variance of $E[X|X + Y_n]$ converges to $\inf_Y \operatorname{var} E[X|X + Y]$. Secondly, by the tightness of $\{Y_n\}_{n=1}^\infty$, there exists a subsequence that converges weakly to a random variable, say, $\tilde{Y}$. It then remains to show that the variance of $E[X|X + \tilde{Y}]$ is exactly $\inf_Y \operatorname{var} E[X|X + Y]$. This represents the most mathematically challenging task of this work. To this end, we develop a convergence result for the variances of conditional expectations (see Theorem 2.2), which is in part inspired by the work of [2].

The remainder of this paper is organized as follows. In Section 3, we prove that the variance of the least favorable noise must be ϵ^2, allowing us to reduce the class of Y when searching for the least favorable noise. Indeed, this result simplifies the three characterization conditions in [4] to two characterization conditions. In Section 4, we consider further properties of $L(X, \epsilon)$. We prove that it is both strictly decreasing and right continuous in ϵ on $[0, \infty)$. Consequently, the maximum of the prediction error

$E[X - E[X|X + Y]]^2$ is strictly increasing as the variance of the noise Y increases. In other words, "more noise makes prediction worse." This conclusively answers Question 4 posed on page 2 of [4].

2 Existence of the least favorable noise

The main result of this section is the proof of the existence of the least favorable noise, which we give in Theorem 2.5.

We begin by recording Lemma 2.1 below. Since this lemma is a direct result of Skorokhod's representation theorem, Fatou's lemma, and Theorem 4.5.2 in [1], we omit the proof.

Lemma 2.1. *Suppose $\{X_n\}_{n=1}^{\infty}$ is a sequence of random variables converging weakly to some random variable X. Further, assume that*

$$\sup_{1 \le n < \infty} E[X_n^2] < \infty.$$

Then

$$E[X^2] \le \liminf_{n \to \infty} E[X_n^2] \ \text{ and } \ \lim_{n \to \infty} E[X_n] = E[X].$$

We proceed to introduce Theorem 2.2 below, which gives a convergence result for the variances of conditional expectations. It is, in part, inspired by the work of [2].

Theorem 2.2. *Let $\{(X_n, Y_n)\}_{n=1}^{\infty}$ be a sequence of random vectors which converges weakly to some random vector (X, Y). If*

$$\sup_{1 \le n < \infty} E[X_n^2] < \infty,$$

then

$$\operatorname{var} E[X|X + Y] \le \liminf_{n \to \infty} \operatorname{var} E[X_n|X_n + Y_n].$$

Proof. We denote $\sup E[X_n^2]$ by M. By the Continuous Mapping Theorem, X_n converges weakly to X. Applying Lemma 2.1 yields

$$E[X^2] \le \liminf E[X_n^2] \le M. \tag{2.1}$$

$$\lim_{n \to \infty} E[X_n] = E[X]. \tag{2.2}$$

By standard properties of conditional expectation, the mean of $E[X_n|X_n + Y_n]$ is $E[X_n]$. Then

$$\operatorname{var} E[X_n|X_n + Y_n] = E\left[(E[X_n|X_n + Y_n])^2\right] - (E[X_n])^2.$$

Similarly, $\operatorname{var} E[X|X + Y] = E\left[(E[X|X + Y])^2\right] - (E[X])^2$. Together with (2.2), we need only show that

$$E\left[(E[X|X + Y])^2\right] \le \liminf_{n \to \infty} E\left[(E[X_n|X_n + Y_n])^2\right]. \tag{2.3}$$

Since (X_n, Y_n) converges weakly to (X, Y), by Skorohod's representation theorem, we can construct (U_n, W_n) and (U, W) on a new probability space such that $(U_n, W_n) \overset{d}{=} (X_n, Y_n)$, $(U, W) \overset{d}{=} (X, Y)$ and (U_n, W_n) converges to (U, W) almost surely. It follows immediately that

$$\sup_{1 \le n < \infty} E[U_n^2] = \sup_{1 \le n < \infty} E[X_n^2] = M,$$

and

$$E[U^2] = E[X^2] \le M.$$

Noting that the distributions of $E[X_n|X_n + Y_n]$ and $E[U_n|U_n + W_n]$ depend only on the distributions, respectively, of (X_n, Y_n) and (U_n, W_n), it follows from the fact

$$(X_n, Y_n) \stackrel{d}{=} (U_n, W_n)$$

that the distribution of $E[X_n|X_n + Y_n]$ coincides with that of $E[U_n|U_n + W_n]$. Thus,

$$E\left[(E[X_n|X_n + Y_n])^2\right] = E\left[(E[U_n|U_n + W_n])^2\right].$$

Similarly,

$$E\left[(E[X|X + Y])^2\right] = E\left[(E[U|U + W])^2\right].$$

Thus, to show (2.3), it suffices to prove

$$E\left[(E[U|U + W])^2\right] \leq \liminf_{n \to \infty} E\left[(E[U_n|U_n + W_n])^2\right]. \tag{2.4}$$

We proceed to define

$$T_n := E[U_n|U_n + W_n] \ \text{ and } \ T := E[U|U + W]. \tag{2.5}$$

By Jensen's inequality, for all $n \in \mathbb{N}_+$,

$$E[T_n^2] = E\left[(E[U_n|U_n + W_n])^2\right] \leq E\left[E[U_n^2|U_n + W_n]\right] = E[U_n^2] \leq M.$$

Similarly, $E[T^2] \leq M$.

The proof continues by invoking the following lemma, whose proof is relegated to the Appendix.

Lemma 2.3. *For every bounded Borel function h on $\mathbb{R}$,*

$$E[h(U + W)\,T] = \lim_{n \to \infty} E[h(U + W)\,T_n]. \tag{2.6}$$

We now show how Lemma 2.3 implies (2.4). By the definition of conditional expectation, $E[U|U + W]$ can be represented as $g(U + W)$, where g is a Borel function on $\mathbb{R}$. For $k \in \mathbb{N}_+$, let $g_k := g\,\mathbb{1}_{\{|g| \leq k\}}$. Then g_k is a bounded Borel function. Applying Lemma 2.3 gives

$$E[g_k(U + W)\,T] = \lim_{n \to \infty} E[g_k(U + W)\,T_n].$$

Since

$$E[g_k(U + W)\,T_n] \leq \left(E\left[(g_k(U + W))^2\right] + E[T_n^2]\right)/2,$$

letting $n \to \infty$ yields

$$E[g_k(U + W)\,T] \leq \frac{1}{2}\liminf_{n \to \infty} E[T_n^2] + \frac{1}{2}\,E\left[(g_k(U + W))^2\right]. \tag{2.7}$$

Noting that

$$g(U + W) = E[U|U + W] = T$$

is square-integrable, it then follows by dominated convergence that

$$\lim_{k \to \infty} E\left[(g_k(U + W))^2\right] = E\left[(g(U + W))^2\right] = E[T^2]. \tag{2.8}$$

Further, noting that

$$|g_k(U + W)\,T| \leq |g(U + W)\,T| = T^2,$$

and recalling that T^2 is integrable, the family of random variables $\{g_k(U + W)\,T\}$ is uniformly integrable. Since $g_k(U + W)\,T$ converges to $g(U + W)\,T$ almost surely, we have that

$$\lim_{k \to \infty} E[g_k(U + W)\,T] = E[g(U + W)\,T] = E[T^2]. \qquad (2.9)$$

Finally, letting $k \to \infty$ on both sides of (2.7), and combining the results in (2.8) and (2.9), we have

$$E[T^2] \le \liminf_{n} E[T_n^2],$$

which is precisely what appears in (2.4). This completes the proof. □

Remark 2.4. If the random vectors (X_n, Y_n) and (X, Y) satisfy the condition that

$$\sup_{B \in \mathcal{B}_{\mathbb{R}}} \left| P\left[(X_n, Y_n) \in A, X_n + Y_n \in B\right] - P\left[(X, Y) \in A, X + Y \in B\right] \right| \to 0$$

for each A in $\mathcal{B}_{\mathbb{R}^2}$, where $\mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\mathbb{R}^2}$ represent the families of Borel sets on $\mathbb{R}$ and $\mathbb{R}^2$ respectively, then Corollary 5.7 in [3] could help with the simplification of the proof for Theorem 2.2. In general, there are situations where this condition is not satisfied, particularly when dealing with jointly discrete random vectors (X_n, Y_n) and (X, Y). Therefore, we provide the above proof to address all situations.

With Theorem 2.2 in hand, we are now ready to present the key theorem of this manuscript regarding existence of the least favorable noise.

Theorem 2.5. *Suppose X is a non-degenerate random variable with zero mean and finite second moment. Then there exists a minimizer $Y \in \mathcal{V}_\epsilon(X)$ such that*

$$\operatorname{var} E[X|X + Y] = \inf_{Z \in \mathcal{V}_\epsilon(X)} \operatorname{var} E[X|X + Z] = L(X, \epsilon).$$

Consequently, the least favorable noise exists.

Proof. Let $Y_n \in \mathcal{V}_\epsilon(X)$ be a random variable such that

$$\operatorname{var} E[X|X + Y_n] \le L(X, \epsilon) + \frac{1}{n}.$$

Since $E[X^2] < \infty$ and $\sup_n E[Y_n^2] \le \epsilon^2$, the sequence of random vectors $\{(X, Y_n)\}_{n=1}^{\infty}$ is tight. Thus, there exists a subsequence $\{(X, Y_{n(k)})\}$ of $\{(X, Y_n)\}$ such that $(X, Y_{n(k)})$ converges weakly to some random vector, say, $(\tilde{X}, \tilde{Y})$. It follows by the Continuous Mapping Theorem that $X \overset{d}{=} \tilde{X}$ and $Y_{n(k)}$ converges weakly to $\tilde{Y}$. Noting that X is independent of Y_n, it may be easily verified that $\tilde{X}$ is independent of $\tilde{Y}$. Furthermore, applying Lemma 2.1 to the sequence $\{Y_{n(k)}\}$, we have

$$E[\tilde{Y}^2] \le \liminf_{k \to \infty} E[Y_{n(k)}^2] \le \epsilon^2 \ \text{ and } \ E[\tilde{Y}] = \lim_{k \to \infty} E[Y_{n(k)}] = 0.$$

We now may construct a random variable Y such that Y is independent of X and Y has the same distribution as $\tilde{Y}$. We now claim Y is the desired minimizer. We proceed to prove this claim. Indeed, since

$$E[Y] = E[\tilde{Y}] = 0$$

and

$$E[Y^2] = E[\tilde{Y}^2] \le \epsilon^2,$$

then $Y \in \mathcal{V}_\epsilon(X)$. Furthermore, it follows by the assumption of independence of X and Y that $(X, Y) \overset{d}{=} (\tilde{X}, \tilde{Y})$. Invoking a similar argument from the proof of Theorem 2.2 yields that

$$\operatorname{var} E[X|X + Y] = \operatorname{var} E[\tilde{X}|\tilde{X} + \tilde{Y}].$$

Finally, applying Theorem 2.2 yields

$$\mathrm{var} E[\tilde{X}|\tilde{X} + \tilde{Y}] \leq \liminf_{k\to\infty} \mathrm{var} E[X_{n(k)}|X_{n(k)} + Y_{n(k)}] = L(X, \epsilon).$$

Thus, $\mathrm{var} E[X|X + Y] \leq L(X, \epsilon)$, which proves that Y is the desired minimizer. $\square$

3 Variance of the least favorable noise

The purpose of this section is to calculate the variance of the least favorable noise (Theorem 3.3). Theorem 3.3 relies on Proposition 3.1 and Lemma 3.2 below. Proposition 3.1 is a standard result conditional expectation and therefore we omit the proof.

Proposition 3.1. *Suppose X is a square-integral random variable. Let $\mathcal{F}$ and $\mathcal{G}$ be two σ-algebras with $\mathcal{G} \subset \mathcal{F}$. Then*

$$E\left[(E[X|\mathcal{G}])^2\right] \leq E\left[(E[X|\mathcal{F}])^2\right],$$

with equality holding if and only if $E[X|\mathcal{G}] = E[X|\mathcal{F}]$ almost surely.

We now introduce Lemma 3.2, which relies on Proposition 3.1.

Lemma 3.2. *Suppose X is a non-degenerate random variable with zero mean and finite second moment and Y is an arbitrary random variable independent of X. Let Z be a non-degenerate Gaussian random variable independent of $\sigma(X, Y)$. Then*

$$\mathrm{var}\, E[X|X + Y + Z] < \mathrm{var}\, E[X|X + Y].$$

Proof. First, we note that the means of $E[X|X + Y]$ and $E[X|X + Y + Z]$ are both $E[X] = 0$. It thus suffices to prove that

$$E\left[(E[X|X + Y + Z])^2\right] < E\left[(E[X|X + Y])^2\right].$$

Applying Proposition 3.1 gives

$$E\left[(E[X|X + Y + Z])^2\right] \leq E\left[(E[X|X + Y, Z])^2\right] = E\left[(E[X|X + Y])^2\right], \qquad (3.1)$$

where in the last equality we have invoked the fact that

$$E[X|X + Y, Z] = E[X|X + Y],$$

which holds because Z is independent of $\sigma(X, Y)$. In what follows, we only need rule out the case where

$$E\left[(E[X|X + Y + Z])^2\right] = E\left[(E[X|X + Y])^2\right]. \qquad (3.2)$$

We proceed by contradiction. Assume, for the sake of contradiction, that equation (3.2) holds. Combining (3.1) and Proposition 3.1, we have that

$$E[X|X + Y + Z] = E[X|X + Y, Z] = E[X|X + Y], \qquad (3.3)$$

holds almost surely.

Let f_1 and f_2 be two Borel functions on $\mathbb{R}$ such that $f_1(X + Y + Z)$ and $f_2(X + Y)$ are versions of $E[X|X + Y + Z]$ and $E[X|X + Y]$ respectively. Then the equality in (3.3) implies that

$$P\left(f_1(X + Y + Z) = f_2(X + Y)\right) = 1. \qquad (3.4)$$

In what follows, we use P_U to denote the probability measure on $\mathbb{R}$ generated by the random variable U. Since X, Y, Z are independent, (3.4) is equivalent to

$$\int_{\mathbb{R}} \int_{\mathbb{R}} P(f_1(x + y + Z) = f_2(x + y))\, P_X(dx)\, P_Y(dy) = 1.$$

Since $P(f_1(x + y + Z) = f_2(x + y)) \leq 1$, we have that

$$P(f_1(x + y + Z) = f_2(x + y)) = 1 \qquad P_X \otimes P_Y \text{ – a.s.}.$$

Then we can select x_0, y_0 such that

$$P(f_1(x_0 + y_0 + Z) = f_2(x_0 + y_0)) = 1,$$

which implies that $f_1(x_0 + y_0 + Z)$ is a constant almost surely. Note that since Z is an absolutely continuous random variable whose density is positive everywhere, f_1 is a constant almost everywhere with respect to Lebesgue measure. By standard properties of convolution, $X + Y + Z$ is an absolutely continuous random variable. Thus, $f_1(X + Y + Z)$ is a constant almost surely, namely, $E[X|X + Y + Z]$ is a constant almost surely. Note that since the mean of $E[X|X + Y + Z]$ is 0, $E[X|X + Y + Z]$ must be 0 almost surely. Then

$$
\begin{aligned}
0 &= E\left[E[X|X + Y + Z](X + Y + Z)\right] = E\left[X(X + Y + Z)\right] \\
&= E[X^2] + E[XY] + E[XZ] = E[X^2] + E[X]E[Y] + E[X]E[Z] = E[X^2],
\end{aligned}
$$

which implies that $X = 0$ almost surely. Since X is non-degenerate, a contradiction has been reached. This concludes the proof. $\qquad\square$

With above lemmas in hand, we now turn to the variance of the least favorable noise.

Theorem 3.3. *Suppose X is a non-degenerate random variable with zero mean and finite second moment. For any $Y \in \mathcal{V}_\epsilon(X)$ such that $\operatorname{var} E[X|X + Y]$ attains the minimum $L(X, \epsilon)$, we must have that $E[Y^2] = \epsilon^2$.*

Proof. We proceed by contradiction. For the sake of contradiction, let us assume $E[Y^2] < \epsilon^2$. We proceed to construct a Gaussian random variable Z with zero mean and sufficiently small variance such that Z is independent of $\sigma(X, Y)$ and $E[(Y + Z)^2] \leq \epsilon^2$. It is immediate to verify that $Y + Z \in \mathcal{V}_\epsilon(X)$. However, applying Lemma 3.2 gives

$$\operatorname{var} E[X|X + Y + Z] < \operatorname{var} E[X|X + Y].$$

However, this directly contradicts the fact that Y is a minimizer. Thus, by contradiction, $E[Y^2] = \epsilon^2$. $\qquad\square$

Remark 3.4. Theorem 3.3 tells us that in order to search for the least favorable noise, one may only need focus on the random variable Y satisfying $E[Y^2] = \epsilon^2$. If one attempts to find the least favorable noise by the method of Lagrange multipliers, Theorem 3.3 converts the inequality constraint to an equality constraint, which greatly simplifies the problem.

4   Properties of $L(X, \epsilon)$

The purpose of this section is to study some properties of the function $L(X, \epsilon)$. Proposition 4.1 is introduced in order to prove the key result of this section (Theorem 4.2), which states that $L(X, \epsilon)$ is strictly decreasing and right continuous in ϵ on $[0, \infty)$.

Proposition 4.1. *Let $\{X_n\}_{n=1}^\infty$ be a sequence of random variables with zero mean which converges weakly to some random variable X with zero mean. Let $\{\epsilon_n\}_{n=1}^\infty$ be a sequence of non-negative real number which converges to a non-negative real number ϵ. If*

$$\sup_{1 \leq n < \infty} E[X_n^2] < \infty,$$

then

$$L(X, \epsilon) \leq \liminf_{n \to \infty} L(X_n, \epsilon_n).$$

Proof. By the properties of limit inferior, there exists a subsequence $\{n(k)\}_{k=1}^{\infty}$ of $\{n\}_{n=1}^{\infty}$ such that

$$\liminf_{n\to\infty} L(X_n, \epsilon_n) = \lim_{k\to\infty} L(X_{n(k)}, \epsilon_{n(k)}).$$

Thus, provided the limit of $L(X_{n(k)}, \epsilon_{n(k)})$ exists, we need only prove

$$L(X, \epsilon) \leq \liminf_{k\to\infty} L(X_{n(k)}, \epsilon_{n(k)}).$$

For simplicity of notation, in what follows, we continue to write the subsequence $\{n(k)\}_{k=1}^{\infty}$ by $\{n\}_{n=1}^{\infty}$. We shall also assume the limit of $L(X_n, \epsilon_n)$ exists.

For every n, by Theorem 2.5, there exists a $Y_n \in \mathcal{V}_{\epsilon_n}(X_n)$ such that

$$\mathrm{var}E[X_n|X_n + Y_n] = L(X_n, \epsilon_n).$$

Noting that $E[Y_n^2] \leq \epsilon_n^2$ and that ϵ_n converges to ϵ, we have $\sup_n E[Y_n^2] < \infty$. Together with the fact that $\sup_n E[X_n^2] < \infty$, the family of random vectors $\{(X_n, Y_n)\}_{n=1}^{\infty}$ is tight. Then there exists a subsequence $\{(X_{n(k)}, Y_{n(k)})\}$ of $\{(X_n, Y_n)\}$ such that $(X_{n(k)}, Y_{n(k)})$ converges weakly to some random vector, say, $(\tilde{X}, \tilde{Y})$. By continuous mapping, we have that $X_{n(k)}$ converges weakly to $\tilde{X}$ and $Y_{n(k)}$ converges weakly to $\tilde{Y}$. Then, $X \stackrel{d}{=} \tilde{X}$, since $X_{n(k)}$ also converges weakly to X. By Lemma 2.1, we have $E[\tilde{Y}] = \lim_k E[Y_{n(k)}] = 0$ and $E[\tilde{Y}^2] \leq \liminf_k E[Y_{n(k)}^2] \leq \epsilon^2$. Furthermore, since $X_{n(k)}$ is independent of $Y_{n(k)}$, we also have that $\tilde{X}$ is independent of $\tilde{Y}$. Invoking Theorem 2.2 here gives

$$\mathrm{var}E[\tilde{X}|\tilde{X} + \tilde{Y}] \leq \liminf_{k\to\infty} \mathrm{var}E[X_{n(k)}|X_{n(k)} + Y_{n(k)}]$$
$$= \liminf_{k\to\infty} L(X_{n(k)}, \epsilon_{n(k)}) = \liminf_{n\to\infty} L(X_n, \epsilon_n), \tag{4.1}$$

where the last equality follows by our assumption that the limit of $L(X_n, \epsilon_n)$ exists.

There exists a random variable Y such that Y is independent of X and $Y \stackrel{d}{=} \tilde{Y}$. Thus, $E[Y] = E[\tilde{Y}] = 0$ and $E[Y^2] = E[\tilde{Y}^2] \leq \epsilon^2$, which implies $Y \in \mathcal{V}_\epsilon(X)$. By independence, we have $(X, Y) \stackrel{d}{=} (\tilde{X}, \tilde{Y})$, hence,

$$\mathrm{var}E[X|X + Y] = \mathrm{var}E[\tilde{X}|\tilde{X} + \tilde{Y}], \tag{4.2}$$

by a similar argument in the proof of Theorem 2.2. Combining (4.1) and (4.2), we have

$$L(X, \epsilon) \leq \mathrm{var}E[X|X + Y] \leq \liminf_{n\to\infty} L(X_n, \epsilon_n).$$

This completes the proof. $\qquad\square$

Theorem 4.2. *Let X be a non-degenerate random variable with zero mean and finite second moment. Then, with fixed X, $L(X, \epsilon)$ is a strictly decreasing and right continuous function with respect to ϵ on $[0, \infty)$.*

Proof. We first shall prove that $L(X, \epsilon)$ is strictly decreasing with respect to ϵ. Consider $0 \leq \epsilon_1 < \epsilon_2$. By Theorem 2.5, there exists $Y_1 \in \mathcal{V}_{\epsilon_1}(X)$ such that

$$\mathrm{var}\, E[X|X + Y_1] = L(X, \epsilon_1).$$

We proceed to construct a Gaussian random variable with mean zero and variance $\epsilon_2^2 - \epsilon_1^2$ such that Z is independent of $\sigma(X, Y_1)$. It is straightforward to check that $Y_1 + Z \in \mathcal{V}_{\epsilon_2}(X)$. Then by Lemma 3.2,

$$L(X, \epsilon_2) \leq \mathrm{var}\, E[X|X + Y_1 + Z] < \mathrm{var}\, E[X|X + Y_1] = L(X, \epsilon_1),$$

which means that $L(X, \epsilon)$ is strictly decreasing.

We now turn to the right continuity. Consider every $\epsilon_0 \in [0, \infty)$. For $\epsilon > \epsilon_0$, we have $L(X, \epsilon) < L(X, \epsilon_0)$. Thus

$$\limsup_{\epsilon \downarrow \epsilon_0} L(X, \epsilon) \leq L(X, \epsilon_0). \tag{4.3}$$

Applying Proposition 4.1 yields

$$L(X, \epsilon_0) \leq \liminf_{\epsilon \downarrow \epsilon_0} L(X, \epsilon). \tag{4.4}$$

Finally, combining (4.3) and (4.4) completes the proof. $\qquad\square$

Remark 4.3. Theorem 4.2 shows that the maximum of the prediction error $E[X - E[X|X + Y]]^2$ is strictly increasing as the variance of the noise Y increases. In other words, "more noise makes prediction worse." This conclusively answers Question 4 posed on page 2 of [4].

References

[1] K.L. Chung, *A course in probability theory*, third ed., Academic Press, Inc., San Diego, CA, 2001. MR1796326

[2] I. Crimaldi and L. Pratelli, *Convergence results for conditional expectations*, Bernoulli **11** (2005), no. 4, 737–745. MR2158258

[3] I. Crimaldi and L. Pratelli, *Two inequalities for conditional expectations and convergence results for filters*, Statist. Probab. Lett. **74** (2005), no. 2, 151–162. MR2169373

[4] P.A. Ernst, A.M. Kagan, and L.C.G.. Rogers, *The least favorable noise*, Electron. Commun. Probab. **27** (2022), Paper No. 28, 11. MR4424034

[5] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 2000, Foundations, Reprint of the second (1994) edition. MR1796539

5 Appendix

This Appendix contains the proof of Lemma 2.3.

Proof of Lemma 2.3. We first claim that the class $\mathcal{H}$ of bounded real-valued Borel function h on $\mathbb{R}$ such that (2.6) holds must satisfy properties (A.1)-(A.3) below:

(A.1) $\mathcal{H}$ is a vector space which contains constant functions;

(A.2) $\mathcal{H}$ is closed under uniform convergence;

(A.3) For a uniformly bounded sequence $\{h_k\}$ of non-negative functions in $\mathcal{H}$ where $\forall k$, $h_k \leq h_{k+1}$, and $\forall s$, $h_k(s) \to h(s)$, we have that $h \in \mathcal{H}$.

It is immediate that $\mathcal{H}$ is a vector space. Furthermore, for every constant function $c(x) \equiv c$, by (2.2), we have

$$\lim_{n \to \infty} E[c(U + W)T_n] = \lim_{n \to \infty} c\,E[T_n] = \lim_{n \to \infty} c\,E[X_n] = c\,E[X] = c\,E[T] = E[c(U + W)T].$$

Thus, property (A.1) holds.

To check property (A.2), suppose $h_k \in \mathcal{H}$ converges uniformly to h. Then

$$
\begin{aligned}
&|E[h(U + W)\, T] - E[h(U + W)\, T_n]| \\
\leq\ & |E\left[(h(U + W) - h_k(U + W))\, T\right]| + |E\left[(h(U + W) - h_k(U + W))\, T_n\right]| \\
& + |E[h_k(U + W)\, T] - E[h_k(U + W)\, T_n]| \\
\leq\ & \left\{ E\left[(h(U+W) - h_k(U+W))^2\right] \right\}^{1/2} \cdot \left\{ E[T^2] \right\}^{1/2} \\
& + \left\{ E\left[(h(U+W) - h_k(U+W))^2\right] \right\}^{1/2} \cdot \left\{ E[T_n^2] \right\}^{1/2} \\
& + |E[h_k(U + W)\, T] - E[h_k(U + W)\, T_n]| \\
\leq\ & 2M^{1/2} \left\{ E\left[(h(U+W) - h_k(U+W))^2\right] \right\}^{1/2} \\
& + |E[h_k(U + W)\, T] - E[h_k(U + W)\, T_n]|,
\end{aligned}
$$

where the second inequality follows by Hölder's inequality and in the last inequality we have applied the fact that $\sup_n E[T_n^2] \leq M$ and $E[T^2] \leq M$. Letting n tend to ∞ and noting that (2.6) holds for h_k, we have

$$
\limsup_{n \to \infty} |E[h(U+W)T] - E[h(U+W)T_n]| \leq 2M^{1/2} \left\{ E\left[(h(U+W) - h_k(U+W))^2\right] \right\}^{1/2}.
\tag{5.1}
$$

Letting k tend to ∞, and recalling the fact that h_k converges uniformly to h, we conclude that $h \in \mathcal{H}$.

We now turn to (A.3). Suppose uniformly bounded non-negative functions $h_k \uparrow h$. It is immediate that $h_k(U + W) - h(U + W)$ converges to 0 pointwise and that it is uniformly bounded. By dominated convergence, we have

$$
\lim_{n \to \infty} E\left[(h(U + W) - h_k(U + W))^2\right] = 0
\tag{5.2}
$$

Using a similar argument in the proof of (A.2) yields (5.1) again. Together with (5.2) we conclude that $h \in \mathcal{H}$.

With above preparation in hand, we prove the equality in (2.6) for every bounded Borel function h. By the monotone class theorem (cf. [5, p.91]), it suffices to prove (2.6) for every bounded continuous function h on $\mathbb{R}$. Recall that by construction, (U_n, W_n) converges to (U, W) almost surely. Since h is continuous, we have $h(U_n + W_n) \to h(U + W)$ almost surely. Since h is bounded, by dominated convergence, we have

$$
\lim_{n \to \infty} E\left[(h(U_n + W_n) - h(U + W))^2\right] = 0.
\tag{5.3}
$$

Note that

$$
\sup_{1 \leq n < \infty} E\left[(U_n h(U_n + W_n))^2\right] \leq \|h\|^2 \cdot \sup_{1 \leq n < \infty} E[U_n^2] = M\|h\|^2,
$$

where $\|h\| := \sup_x |h(x)|$. Then the family of random variables $\{U_n h(U_n + W_n)\}_{n=1}^\infty$ is uniformly integrable. Further, $U_n h(U_n + W_n)$ converges to $U h(U + W)$ almost surely, and so

$$
\lim_{n \to \infty} E[U_n h(U_n + W_n)] = E[U h(U + W)].
\tag{5.4}
$$

Recalling the definitions of T_n and T given in (2.5), and invoking the standard properties of conditional expectation, we have

$$
E[h(U_n + W_n)\, T_n] = E\left[h(U_n + W_n)E[U_n | U_n + W_n]\right] = E[U_n h(U_n + W_n)],
$$

and

$$
E[h(U + W)\, T] = E\left[h(U + W)E[U | U + W]\right] = E[U h(U + W)].
$$

Plugging the above two displays into (5.4) yields

$$\lim_{n \to \infty} E[h(U_n + W_n) T_n] = E[h(U + W) T].\tag{5.5}$$

By Hölder's inequality, we have

$$
\begin{aligned}
&\quad\; |E[h(U_n + W_n) T_n] - E[h(U + W) T_n]| \\
&= |E[(h(U_n + W_n) - h(U + W)) T_n]| \\
&\leq \left\{ E\left[(h(U_n + W_n) - h(U + W))^2 \right] \right\}^{1/2} \cdot \left\{ E[T_n^2] \right\}^{1/2} \\
&\leq M^{1/2} \left\{ E\left[(h(U_n + W_n) - h(U + W))^2 \right] \right\}^{1/2}.
\end{aligned}
$$

Together with (5.3), we obtain

$$\lim_{n \to \infty} |E[h(U_n + W_n) T_n] - E[h(U + W) T_n]| = 0\tag{5.6}$$

Combining (5.5) and (5.6), the equality in (2.6) follows. $\qquad\square$

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System: https://vtex.lt/services/ejms-peer-review/
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: https://imstat.org/shop/donation/