A generalization of martingale theory to self-averaging processes

Bo Wang
Charlottesville, VA

B.Sc, Xiamen University, 2012

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Mathematics

University of Virginia
May, 2017

Abstract

We introduce and study a generalization of martingales with the following self-averaging property: at each time, the conditional expectation of future random variables given the past, is a weighted average of all the random variables comprising the past. We assume only that more recent random variables are weighted no less than older random variables. We investigate conditions under which important properties satisfied by martingales, such as maximal inequalities and convergence, are present in an appropriate form.

Contents

Chapter 1

Background

A *martingale* is a sequence of integrable random variables $\{X_n\}$ with the property that, conditional on knowing the values of the first n variables $X_1, \ldots, X_n$, the expected value of X_{n+1} equals the most recent known value X_n. That is, for each $n = 1, 2, \ldots,$

$$E[X_{n+1} | \mathcal{F}_n] = X_n,$$

where $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is the σ-algebra generated by the first n random variables. Recall that in the above display, the *conditional expectation* $E[X_{n+1}|\mathcal{F}_n]$ is the unique (up to null sets) random variable Y such that $E[Y 1_A] = E[X_{n+1} 1_A]$ for all $A \in \mathcal{F}_n$, where 1_A denotes the indicator function on the event A. Informally, $E[X_{n+1}|\mathcal{F}_n]$ is the best guess for X_{n+1} given the information $\mathcal{F}_n$, which in this case consists of the values of $X_1, \ldots, X_n$. So if one observes $5, 2, -11, 7, 9$ as the first five values of a martingale, the best guess for X_6 would be 9.

A basic example of a martingale is *simple symmetric random walk*: let $\{\xi_i\}$ be an independent and identically distributed (iid) sequence of random variables with

$P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, and define the random walk

$$S_n = \sum_{i=1}^{n} \xi_i, \qquad (n \geq 1),$$

the location at time n of the random walker who, independently at each step, takes a step up or down with equal probability. Since $S_{n+1} = S_n + \xi_{n+1}$, and since S_n is $\mathcal{F}_n$-measurable but ξ_{n+1} is independent of $\mathcal{F}_n$ with mean zero, we easily verify that

$$E[S_{n+1}|\mathcal{F}_n] = S_n + E[\xi_{n+1}] = S_n,$$

and so $\{S_n\}$ is a martingale. This is intuitive – since the walker steps up or down with equal probability, the expected value of his next location is simply his current position.

Martingales were introduced in 18th century France to study the properties of certain gambling strategies popular at that time. In fact the term martingale is believed to derive from a gambling strategy of the same name, a strategy so absurd, that gamblers employing it were said to be playing like Martigals (*"jouga a la martegalo"*), residents of the Provençal town of Martigues, who had a reputation for naïveté or foolishness [4].

The martingale strategy is played against a fair game, such as coin flips, in which you lose your bet on tails and double it on heads. To play, you first bet one dollar. If you win, you walk away with an extra dollar; if you lose you play again, betting double your last bet, or two dollars. If you win the second game, you are one dollar ahead and walk away. If not you play again, doubling down with four dollars. Proceeding in this way, you bet 2^{k-1} on the kth game until you win, which on the nth game yields winnings of $2^{n-1} = \sum_{k=1}^{n-1} 2^{k-1} + 1$, recovering all of your losses and netting one dollar profit. Moreover, since the probability of losing forever is $\lim_{n\to\infty} 1/2^n = 0$, you are guaranteed with probability one to win in finite time and make a \$1 profit each time you play this way.

This situation can be described mathematically as follows. If $\{\xi_i\}$ are as above and represent the sequence of coin flips (1 for heads, −1 for tails), we can define $B_1 = 1$ and for $k \geq 2$ the kth bet as

$$B_k = \begin{cases} 2B_{k-1}, & \xi_{k-1} = -1, \\ 0, & \xi_{k-1} = 1. \end{cases}$$

Notice that the B_k are eventually all zero after the first time a ξ_k equals one, meaning that betting stops at that point. The total profit after n games is given by

$$M_n = \sum_{k=1}^{n} B_k \xi_k.$$

Letting $\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n)$ we see that B_{n+1} as well as M_n are $\mathcal{F}_n$-measurable and ξ_{n+1} is independent of $\mathcal{F}_n$ with mean zero. So since $M_{n+1} = M_n + B_{n+1}\xi_{n+1}$, we have

$$E[M_{n+1}|\mathcal{F}_n] = M_n + E[B_{n+1}\xi_{n+1}|\mathcal{F}_n] = M_n + B_{n+1}E[\xi_{n+1}] = M_n,$$

and so $\{M_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$.

In fact it is the "original" martingale and we have already seen that almost surely (i.e. with probability one) $M_n \to 1$ in finite time as $n \to \infty$. Consequently, playing "martingale" seemed a sure thing in 18th century gambling circles, as long as you had the courage to wait for that first win. On the other hand, experience showed more than the occasional gambler ruined while trying, so many considered it a strategy of fools.

Even then it must have been clear that some aspects of the strategy are unrealistic: casinos typically limit the maximum bet, no gambler can play for arbitrary lengths of time, and most importantly, no gambler can fund arbitrarily high bets. Nevertheless these restrictions may seem unimportant, as they are only relevant on events with very small probability. It took some time for the development of the mathematical theory of martingales to fully explain these seeming paradoxes.

To complete the above story, to motivate the importance of martingale theory, and to motivate the generalization that is the main topic of this work, we now briefly review a few of the most famous results concerning martingales, which are due to J.L. Doob [1]. An excellent introduction to the theory can be found in [7].

We first provide a slightly more formal definition than was given above.

Definition 1.0.1. *A* filtration $\{\mathcal{F}_n\}$ *on a probability space* $(\Omega, \mathcal{F}, P)$ *is a sequence of increasing sub-σ-algebras of* $\mathcal{F}$*,*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}.$$

A sequence of integrable random variables $\{X_n\}$ *on* $(\Omega, \mathcal{F}, P)$ *is called a* submartingale *with respect to* $\{\mathcal{F}_n\}$ *if for each* n, X_n *is* $\mathcal{F}_n$*-measurable and*

$$E[X_{n+1}|\mathcal{F}_n] \geq X_n. \tag{1.0.1}$$

If the inequality is reversed, then $\{X_n\}$ *is a* supermartingale*. A sequence that is both a sub- and supermartingale (that is, with the inequality replaced by equality) is a* martingale*.*

Sub(super)martingales are weaker notions than that of a martingale in that they only require the expected future to be above(below) the current value of the process. Submartingales drift upwards; supermartingales drift downwards, and martingales are in a sense constant on average. Note that for a martingale (i.e. using (1.0.1) with equality), the tower property of conditional expectation immediately implies that for all $n \geq m$

$$E[X_n|\mathcal{F}_m] = X_m. \tag{1.0.2}$$

That is, the expected value of any future random variable, given the information $\mathcal{F}_m$, still equals the most recent known variable X_m. A martingale is called bounded in L^p if $\sup_n \|X_n\|_p < \infty$. We define the running maximum X_n^* at time n as $X_n^* = \max_{1\leq k\leq n} X_k$.

The first result we highlight is in a sense a generalization of Chebyshev's inequality to the running maximum of a martingale.

Theorem 1.0.2. *(Doob's maximal inequality) If* $\{X_n\}$ *is a non-negative martingale, then for all* n *and* $a > 0$*,*

$$aP(X_n^* \geq a) \leq E[X_n; X_n^* \geq a] \leq E[X_n].$$

The nice thing about this theorem is that it not only bounds the probability in terms of an expectation, but the expectation is just of the terminal random variable X_n (and thus easier to compute), as opposed to of the running maximum as Chebyshev would give.

There is also an L^p version of this inequality.

Theorem 1.0.3. *(Doob's L^p inequality) If $\{X_n\}$ is a non-negative martingale bounded in L^p for $p > 1$, then for all n,*

$$\|X_n^*\|_p \le \frac{p}{p-1}\|X_n\|_p.$$

The following two results are perhaps the most famous, and are the key to finishing the gambling story above.

Theorem 1.0.4. *(Doob's convergence theorem) Let $\{X_n\}$ be a submartingale whose positive part $X_n^+ = X_n \vee 0$ is bounded in L^1. Then almost surely $\lim_{n\to\infty} X_n$ exists and is finite. Moreover, if $\{X_n\}$ are uniformly integrable, that is*

$$\lim_{C\to\infty} \sup_n E[|X_n|; |X_n| > C] = 0,$$

then there exists an integrable random variable X such that $X_n \to X$ a.s. and in L^1.

There is also the L^2 version.

Theorem 1.0.5. *(L^2 convergence) Let $\{X_n\}$ be a martingale and define $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$. Then $\{X_n\}$ is bounded in L^2 if and only if $\sum E[\epsilon_n^2] < \infty$. If either (and thus both) of these conditions holds, then X_n converges almost surely and in L^2.*

A way to interpret Theorem 1.0.4 is as a stochastic analogue to the following well-known fact: a monotone nondecreasing sequence of real numbers that is bounded above must converge to a finite limit. Indeed the above theorem states that a sequence of random variables that is "monotone nondecreasing on average," in the sense of (1.0.1), and is "bounded above on average" must converge to a finite random limit. Of course the behavior is somewhat richer and the mode of convergence can vary depending on the aforementioned uniform integrability condition.

Thus we see that sub- and supermartingales are simply the stochastic analogues of the well-known monotone sequences. And so martingales are the analogues of the constant sequences. Just as monotone sequences play a very basic role in the proofs of mathematical statements in a deterministic setting, martingales (and sub/supermartingales) play the analogous role in the proofs of myriad results in random settings. That is, although they were historically introduced in the context described earlier, their properties are so fundamental to describing relationships among random variables that they have become basic building blocks of modern probability theory.

Theorem 1.0.4 provides the explanation for the properties of the martingale betting strategy. Note that M_n is a martingale (so in particular a submartingale) that is bounded above by 1. So the mean of its positive part is also bounded by 1 and thus it must converge a.s. to a finite limit. Indeed we already could see that $M_n \to 1$ a.s. as $n \to \infty$. However, because $M_n = -\sum_{k=1}^{n} 2^{k-1}$ with probability $1/2^n$, we see that $\{M_n\}$ are not uniformly integrable, and so they do not converge in mean to 1. In fact $E[M_n] = 0$ for all n, as is easily seen. So while a majority of gamblers playing martingale will walk away with \$1, these will be balanced on average by a number of gamblers who lose big for long periods of time.

It gets worse. As mentioned earlier, no gambler has the funds to actually play martingale. A gambler who thinks he will play martingale is actually playing a slight modification in which the earnings $\tilde{M}_n$ are bounded below by $-C$ (the maximum allowable debt). Since $\{\tilde{M}_n\}$ is uniformly integrable, by the martingale convergence theorem $\tilde{M}_n$ will actually converge a.s. and in mean to a limiting random variable equal to 1 with high probability and equal to $-C$ with small positive probability. This makes explicit the risk and reward faced by any player with finite funds, and shows that it is not true that a gambler will always reach a profit of \$1.

Chapter 2

Generalized Self-Averaging Processes

In this chapter we introduce the main object of study for this work, which is a new and substantial generalization of the concept of a martingale. Our primary motivation for this stems not from a particular application, but rather from the observation that the martingale property is a particular way of imposing a simple structure on a collection of random variables (in particular a monotone structure) that despite its simplicity can be used to prove many additional properties.

To expand on this observation a little, consider that a martingale $\{X_n\}$ is a collection of possibly very correlated random variables. It is exactly the complex correlation structure of families of random variables that inserts difficulty into probability theory. The martingale property is a way to characterize a correlation structure through a family of conditional expectations of a variable given the values of some of the others. Specifically, the structure is imposed by requiring that

$$E[X_{n+1}|\sigma(X_1,\dots,X_n)] = f_n(X_1,\dots,X_n), \qquad (n \geq 1),$$

where for a martingale, the function $f_n(x_1,\dots,x_n) = x_n$.

Since this formulation proves so useful, a natural question is whether one can effectively study other dependence structures by choosing different functions f_n to describe them. This is the question taken up here, where we consider a family of functions f_n that gives rise to processes that are more general than martingales, but still include martingales as a special case and still exhibit some analogous behaviors.

We consider functions of the from $f(x_n,\dots,x_1) = \tau_n x_n + \dots + \tau_1 x_1$ where the right side is a convex combination such that $\tau_{k+1} \geq \tau_k$. This defines a process $\{X_n\}$ with the property that the conditional expectation of X_{n+1} given the past is a decreasing weighted average of the entire past. Here decreasing means that variables further into the past cannot have more weight than more recent variables. Rather than exhibiting a monotone structure, these processes have a self-averaging structure and are defined formally below.

Although our primary motivation is to explore the extent to which the above perspective generalizes, we note that there are already interesting examples that fall within the scope of our definition and thus provide a supply of applications for our results. For example, consider modeling an individual's credit score $X_1, X_2, \dots$ as it evolves in time. Imagine the customer was very unreliable 5 years ago but quite reliable recently. One might like a model in which his next credit score, although random, is highly correlated to his recent scores, but also retains some dependency on the old, poor score, even if that dependency is weak. One can easily define such a model that fits neatly into the class considered here, while clearly not being a martingale. Some other simple examples will be described further down.

We also note that there have been other generalizations of martingales proposed. Hammersley [3] considers the natural extension of the martingale property to collections of random variables indexed by the integer lattice, which he calls a harness. There is also the notion of quasi-martingale introduced by Fisk, see [2]. Other generalizations such as amart, martingale in the limit, game fairer with time, progressive martingale, and eventual martingale have been defined as well; a nice review and analysis of some properties appears in Tomkins [6]. None of these generalizations allow for the long term dependence considered here.

We now formally define our objects of study.

Definition 2.0.1 (Self-averaging Triangular Array)**.**

Let T be a triangular array of real numbers

$$T = \begin{cases} \tau_1^{2,1} & & & & \\ \tau_2^{3,2} & \tau_1^{3,2} & & & \\ \vdots & \vdots & \ddots & & \\ \tau_{n-1}^{n,n-1} & \tau_{n-2}^{n,n-1} & \cdots & \tau_1^{n,n-1} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{cases} \tag{2.0.1}$$

We call T self-averaging if for n ≥ 2,

$$\tau_{n-1}^{n,n-1} \geq \tau_{n-2}^{n,n-1} \geq \cdots \geq \tau_{1}^{n,n-1} \geq 0,$$

and

$$\tau_{n-1}^{n,n-1} + \tau_{n-2}^{n,n-1} + \cdots + \tau_{1}^{n,n-1} = 1.$$

A good example to keep in mind is to first choose a sequence of non-negative numbers $\{c_n\}_{n\geq 1}$ such that for all n, $c_n \geq c_{n+1}$, and then let

$$\tau_{n}^{n+1,n} = \frac{c_1}{\sum_{i=1}^{n} c_i}, \cdots, \tau_{1}^{n+1,n} = \frac{c_n}{\sum_{i=1}^{n} c_i}. \tag{2.0.2}$$

Definition 2.0.2 (Generalized self-averaging process)**.**

Let T be a self-averaging triangular array. We say a process $\{X_n\}_{n\geq 1}$ is a generalized self-averaging process (GSAP) with respect to T if for n ≥ 2

$$E[X_n|\mathcal{F}_{n-1}] = \tau_{n-1}^{n,n-1} X_{n-1} + \tau_{n-2}^{n,n-1} X_{n-2} + \cdots + \tau_{1}^{n,n-1} X_1. \tag{2.0.3}$$

where $\{\mathcal{F}_n\}_{n\geq 1}$ is a filtration.

Observe that if the first column of the self-averaging array contains all 1's and all other coluns are zero,

$$T = \begin{cases} 1 & & & & \\ 1 & 0 & & & \\ \vdots & \vdots & \ddots & & \\ 1 & 0 & \cdots & 0 & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{cases}$$

then (2.0.3) defines a martingale. Thus GSAP is a generalization of the classic martingale process. Also note that if we choose $\{c_n\} = \{1, 0, 0, \cdots\}$, we can use (2.0.2) to build this T.

A few other simple examples are useful to keep in mind.

Example 2.0.1. *If we choose the sequence* $\{c_n\} = \{1, 1, 0, 0, 0, \cdots\}$, *then using* (2.0.2),

$$T = \begin{cases} 1 \\ 1/2 & 1/2 \\ 1/2 & 1/2 & 0 \\ \vdots & \vdots & \vdots & \ddots \\ 1/2 & 1/2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{cases}$$

Then for $n \geq 3$, *the conditional expectation is the average of the previous two terms:*

$$E[X_n|\mathcal{F}_{n-1}] = \frac{1}{2}X_{n-1} + \frac{1}{2}X_{n-2}.$$

Example 2.0.2. *If we choose the sequence* $\{c_n\} = \{1, 1, 1, 1, \cdots\}$, *then by* (2.0.2),

$$T = \begin{cases} 1 \\ 1/2 & 1/2 \\ \vdots & \vdots & \ddots \\ 1/n & 1/n & \cdots & 1/n \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{cases}$$

Then for $n \geq 3$, the conditional expectation is the average of the entire history:

$$E[X_n|\mathcal{F}_{n-1}] = \frac{1}{n-1}X_{n-1} + \frac{1}{n-1}X_{n-2} + \cdots + \frac{1}{n-1}X_1.$$

Example 2.0.3. *If we use a general non-negative non-increasing sequence $\{c_n\}$, then using* (2.0.2) *and the definition of GSAP, we have*

$$E[X_n|\mathcal{F}_{n-1}] = \frac{c_1}{\sum_{i=1}^n c_i}X_{n-1} + \frac{c_2}{\sum_{i=1}^n c_i}X_{n-2} + \cdots + \frac{c_n}{\sum_{i=1}^n c_i}X_1.$$

At first glance, especially in light of Example 2.0.1, the GSAP may appear similar to the well-known autoregressive (AR) models commonly used in time series analysis and signal processing. An autoregressive model of order p is defined as

$$X_n = \phi_1 X_{n-1} + \phi_2 X_{n-2} + \cdots + \phi_p X_{n-p} + e_n$$

where $\{e_n\}$ are i.i.d mean 0 random variables. When compared to the AR model, we can see GSAP as a generalization in one sense and specialization in another. The generalization comes from the fact that in our model the recursion formula allows a different set of weights to be used in each step. Moreover, the AR model uses a fixed history window of p terms while a GSAP can depend on the entire history as in Example 2.0.2. One might think of a GSAP as an AR model of infinite order in some sense, although the weights need to also change at each step in this case. The specialization, on the other hand, comes from the non-negativity of the τ's, while in the AR model, the coefficients can be positive or negative.

Chapter 3

Basic Properties of GSAPs

In this chapter, we will establish some basic properties of GSAPs that are important for their analysis.

The first one is in analogy to the property (1.0.2) for martingales, and states that $E[X_n|\mathcal{F}_m]$ is still a decreasing weighted average of the history up to time m. It also follows from the tower property for conditional expectation.

Lemma 3.0.1. *Let $\{X_n\}$ be a GSAP for the array T. Then for all $n \geq m \geq 1$, there exists constants $\tau_m^{n,m}, \cdots, \tau_1^{n,m}$ such that*

$$E[X_n|\mathcal{F}_m] = \tau_m^{n,m} X_m + \cdots + \tau_1^{n,m} X_1.$$

In particular, $\tau_n^{n,n} = 1$, $\tau_k^{n,n} = 0$ for $k = 1, 2, \cdots, n-1$. Moreover, $\{\tau_k^{n,m}\}$ satisfy

$$\tau_m^{n,m} \geq \cdots \geq \tau_1^{n,m} \geq 0,$$

and

$$\tau_m^{n,m} + \cdots + \tau_1^{n,m} = 1,$$

and the relations

$$\tau_k^{n,m-1} = \tau_k^{n,m} + \tau_m^{n,m} \tau_k^{m,m-1}, \tag{3.0.1}$$

for $k = 1, 2, \cdots, n-1$.

Proof. We proceed by induction. When $m = n$, this is trivial and when $m = n-1$, this is just (2.0.3). Suppose for some $m < n$ we have

$$E[X_n|\mathcal{F}_m] = \tau_m^{n,m} X_m + \cdots + \tau_1^{n,m} X_1$$

with $\tau_m^{n,m} \geq \cdots \geq \tau_1^{n,m} \geq 0$ and $\tau_m^{n,m} + \cdots + \tau_1^{n,m} = 1$.

Then by the tower property,

$$E[X_n|\mathcal{F}_{m-1}] = E[E[X_n|\mathcal{F}_m]|\mathcal{F}_{m-1}] = E[\tau_m^{n,m} X_m + \cdots + \tau_1^{n,m} X_1|\mathcal{F}_{m-1}]$$

Since all but the first term on the right is $\mathcal{F}_{m-1}$-measurable,

$$E[X_n|\mathcal{F}_{m-1}] = \tau_m^{n,m} E[X_m|\mathcal{F}_{m-1}] + \tau_{m-1}^{n,m} X_{m-1} + \cdots + \tau_1^{n,m} X_1.$$

We can again apply (2.0.3) to get

$$\begin{aligned} E[X_n|\mathcal{F}_{m-1}] &= \left(\tau_{m-1}^{n,m} + \tau_m^{n,m}\tau_{m-1}^{m,m-1}\right) X_{m-1} + \cdots + \left(\tau_1^{n,m} + \tau_m^{n,m}\tau_1^{m,m-1}\right) X_1 \\ &= \tau_{m-1}^{n,m-1} X_{m-1} + \cdots + \tau_1^{n,m-1} X_1 \end{aligned}$$

where the second equality serves as the definition of the constants $\tau_{m-1}^{n,m-1}, \cdots, \tau_1^{n,m-1}$. It is easy to see that

$$\tau_{m-1}^{n,m-1} \geq \cdots \geq \tau_1^{n,m-1} \geq 0$$

since $\tau_m^{n,m} \geq \tau_{m-1}^{n,m} \geq \cdots \geq \tau_1^{n,m} \geq 0$, and $\tau_{m-1}^{m,m-1} \geq \cdots \geq \tau_1^{m,m-1}$. Also we have

$$\begin{aligned} \tau_{m-1}^{n,m-1} + \cdots + \tau_1^{n,m-1} &= \left(\tau_{m-1}^{n,m} + \tau_m^{n,m}\tau_{m-1}^{m,m-1}\right) + \cdots + \left(\tau_1^{n,m} + \tau_m^{n,m}\tau_1^{m,m-1}\right) \\ &= \tau_{m-1}^{n,m} + \cdots + \tau_1^{n,m} + \tau_m^{n,m}\left(\tau_{m-1}^{m,m-1} + \cdots + \tau_1^{m,m-1}\right) \\ &= \tau_{m-1}^{n,m} + \cdots + \tau_1^{n,m} + \tau_m^{n,m} = 1 \end{aligned}$$

□

If instead of the previous proof we use forward induction on n, we obtain an alternative recurrence formula for $\tau_k^{n,m}$.

Lemma 3.0.2. *For $m < n$ and $k = 1, 2, \cdots, m$, $\tau_k^{n,m}$ is uniquely determined by $\tau_l^{n,n-1}$ and $\tau_k^{j,m}$ for $l = 1, 2, \cdots, n-1$ and $j = m+1, m+2, \cdots, n-1$.*

Proof. We have

$$E[X_n|\mathcal{F}_m] = E[E[X_n|\mathcal{F}_{n-1}]|\mathcal{F}_m] = E[\tau_{n-1}^{n,n-1}X_{n-1} + \cdots + \tau_1^{n,n-1}X_1|\mathcal{F}_m]$$

Since the last m terms are $\mathcal{F}_m$-measurable,

$$\begin{aligned} E[X_n|\mathcal{F}_m] &= \tau_{n-1}^{n,n-1}E[X_{n-1}|\mathcal{F}_m] + \cdots + \tau_{m+1}^{n,n-1}E[X_{m+1}|\mathcal{F}_m] + \tau_m^{n,n-1}X_m + \cdots + \tau_1^{n,n-1}X_1 \\ &= \tau_{n-1}^{n,n-1}\sum_{i=1}^{m}\tau_i^{n-1,m}X_i + \cdots + \tau_{m+1}^{n,n-1}\sum_{i=1}^{m}\tau_i^{m+1,m}X_i + \tau_m^{n,n-1}X_m + \cdots + \tau_1^{n,n-1}X_1 \\ &= \left(\tau_{n-1}^{n,n-1}\tau_m^{n-1,m} + \cdots + \tau_{m+1}^{n,n-1}\tau_m^{m+1,m} + \tau_m^{n,n-1}\right)X_m \\ &\quad + \cdots + \left(\tau_{n-1}^{n,n-1}\tau_1^{n-1,m} + \cdots + \tau_{m+1}^{n,n-1}\tau_1^{m+1,m} + \tau_1^{n,n-1}\right)X_1 \end{aligned}$$

So for $k \le m < n$,

$$\tau_k^{n,m} = \tau_{n-1}^{n,n-1}\tau_k^{n-1,m} + \cdots + \tau_{m+1}^{n,n-1}\tau_k^{m+1,m} + \tau_k^{n,n-1}$$

□

We can represent the coefficients given in Lemma 3.0.1 using matrix operations.

For each $n \geq 1$, define a family of $n \times n$ matrices $\{M_k : k = 2, \cdots, n\}$ by

$$M_k = \begin{pmatrix} 1 & & & \tau_1^{k,k-1} & & & \\ & \ddots & & \vdots & & & \\ & & 1 & \tau_{k-1}^{k,k-1} & & & \\ & & & 0 & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where only the k^{th} column is different from that of an identity matrix. Then M_k is the operation that splits X_k according to the conditional expectation with respect to $\mathcal{F}_{k-1}$. That is, $E[X_n|\mathcal{F}_m]$ is a weighted average of $X_1, X_2, \cdots, X_m$ with weights given by the first m coordinates of $M_{m+1}M_{m+2}\cdots M_n e_n$

Note the family of matrices have three properties coming from the underlying conditional expectations:

1. Each M_k is idempotent.
2. The sum of every column is 1. This is also true for arbitrary products of the matrices.
3. For $i < j$: $M_i M_j = M_i M_j M_i$.

Chapter 4

Constructing GSAPs

A natural question is how to construct a GSAP. Similar to martingales, they can be constructed as sums of independent mean-zero random variables like in (1.0.1).

4.1 Sums of independent random variables

We first show that any GSAP has a representation as a weighted sum of mean-zero random variables.

Theorem 4.1.1. *If $\{X_n\}$ is a GSAP, then for all $n \geq 2$,*

$$X_n = \epsilon_n + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \cdots + \tau_1^{n,1}\epsilon_1 + \mu \tag{4.1.1}$$

where ϵ_i is $\mathcal{F}_i$-measurable such that $E[\epsilon_i|\mathcal{F}_{i-1}] = 0$ and $\mu = E[X_1]$. The $\tau_k^{n,m}$ are as in (3.0.1)*.*

Proof. Take $\mu = E[X_1]$ and $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$. Then

$$X_n = E[X_n|\mathcal{F}_{n-1}] + \epsilon_n = \tau_{n-1}^{n,n-1}X_{n-1} + \tau_{n-2}^{n,n-1}X_{n-2} + \cdots + \tau_1^{n,n-1}X_1 + \epsilon_n$$

Replacing $X_{n-1} = E[X_{n-1}|\mathcal{F}_{n-2}] + \epsilon_{n-1}$ gives

$$
\begin{aligned}
X_n &= \left(\tau_{n-2}^{n,n-1} + \tau_{n-1}^{n,n-1}\tau_{n-2}^{n-1,n-2}\right) X_{n-2} + \cdots \\
&\quad + \left(\tau_1^{n,n-1} + \tau_{n-1}^{n,n-1}\tau_1^{n-1,n-2}\right) X_1 + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \epsilon_n \\
&= \tau_{n-2}^{n,n-2} X_{n-2} + \cdots + \tau_1^{n,n-2} X_1 + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \epsilon_n
\end{aligned}
$$

Iterating this process gives the result. □

Corollary 4.1.2. *For $m < n$ and ϵ_i as in the previous theorem,*

$$E[X_n|\mathcal{F}_m] = \tau_m^{n,m}\epsilon_m + \cdots + \tau_1^{n,1}\epsilon_1 + \mu.$$

Proof. We have $X_n = \epsilon_n + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \cdots + \tau_1^{n,1}\epsilon_1 + \mu$. Note that for $m + 1 \le i \le n$, $\epsilon_i = X_i - E[X_i|\mathcal{F}_{i-1}]$ satisfies $E[\epsilon_i|\mathcal{F}_{i-1}] = 0$ and hence $E[\epsilon_i|\mathcal{F}_m] = 0$.

So

$$E[\epsilon_n + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \cdots + \tau_{m+1}^{n,m+1}\epsilon_{m+1}|\mathcal{F}_m] = 0$$

and this gives the result. □

We can also prove the converse of Theorem 4.1.1.

Theorem 4.1.3. *Let the triangular array T be self-averaging, and suppose $\{\epsilon_i\}$ are random variables such that $E[\epsilon_i|\mathcal{F}_{i-1}] = 0$ where $\mathcal{F}_i = \sigma(\epsilon_1, \cdots, \epsilon_i)$. Then*

$$X_n = \epsilon_n + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \cdots + \tau_1^{n,1}\epsilon_1 + \mu \tag{4.1.2}$$

is a GSAP, where $\tau_m^{n,m}$ are the coefficients associated with T given by (3.0.1).

Proof. Consider the vector space generated by $\{\epsilon_1, \cdots, \epsilon_{n-1}\}$. Then it is easy to see $\{X_1 - \mu, \cdots, X_{n-1} - \mu\}$ generates the same vector space. If we use $\{\epsilon_1, \cdots, \epsilon_{n-1}\}$ as a basis, then by definition of $\{X_n\}$,

$E[X_n|\mathcal{F}_{n-1}] - \mu$ has coordinates $\left(\tau_1^{n,1}, \tau_2^{n,2}, \cdots, \tau_{n-2}^{n,n-2}, \tau_{n-1}^{n,n-1}\right)^T$,

$X_{n-1} - \mu$ has coordinates $\left(\tau_1^{n-1,1}, \tau_2^{n-1,2}, \cdots, \tau_{n-2}^{n-1,n-2}, \tau_{n-1}^{n-1,n-1}\right)^T$,

$X_{n-2} - \mu$ has coordinates $\left(\tau_1^{n-2,1}, \tau_2^{n-2,2}, \cdots, \tau_{n-2}^{n-2,n-2}, 0\right)^T$,

$\cdots$

$X_1 - \mu$ has coordinates $\left(\tau_1^{1,1}, 0, \cdots, 0\right)^T$.

Then since $\{X_{n-1} - \mu, \cdots, X_1 - \mu\}$ is another basis, we can apply the change of basis formula. So $E[X_n|\mathcal{F}_{n-1}]-\mu$ has coordinates $\left(\tau_{n-1}^{n,n-1}, \cdots, \tau_1^{n,n-1}\right)^T$ under the basis $\{X_{n-1}-\mu, \cdots, X_1 - \mu\}$ given by

$$\begin{pmatrix} \tau_1^{n,1} \\ \vdots \\ \tau_{n-2}^{n,n-2} \\ \tau_{n-1}^{n,n-1} \end{pmatrix} = \begin{pmatrix} \tau_1^{n-1,1} & \tau_1^{n-2,1} & \cdots & \tau_1^{2,1} & 1 \\ \tau_2^{n-1,2} & \tau_2^{n-2,2} & \cdots & 1 & \\ \vdots & \vdots & \iddots & & \\ \tau_{n-2}^{n-1,n-2} & 1 & & & \\ 1 & & & & \end{pmatrix} \begin{pmatrix} \tau_{n-1}^{n,n-1} \\ \vdots \\ \tau_2^{n,n-1} \\ \tau_1^{n,n-1} \end{pmatrix} \tag{4.1.3}$$

□

Note that in the above theorem, we require specific coefficients $\tau_m^{n,m}$ associated with the self-averaging array T. We have to calculate them to construct the GSAP. The change of basis matrix on the right hand side of (4.1.3) depends on the array T up to the $(n-1)^{st}$ row.

4.2 Constructing processes that are almost GSAP

As Theorem 4.1.3 makes clear, we have to calculate the coefficients $\tau_m^{n,m}$ to construct a GSAP. And that requires a fair amount of computation. Here we will give a construction of a process that is much easier to compute, but it is not quite a GSAP. However, it yields a process that behaves almost like a GSAP.

Suppose we have a sequence of numbers $\{a_0 = 1, a_1, a_2, \cdots\}$ and a sequence of random variables $\{\epsilon_i\}$, satisfying $E[\epsilon_i|\mathcal{F}_{i-1}] = 0$, where $\mathcal{F}_i = \sigma(\epsilon_1, \cdots, \epsilon_i)$. For all n, define

$$X_n = a_0\epsilon_n + a_1\epsilon_{n-1} + \cdots + a_{n-1}\epsilon_1. \tag{4.2.1}$$

If we posit that $\{X_n\}$ might behave somewhat like a GSAP, we can change coordinates to write the $\{\epsilon_n\}$ in terms of the $\{X_n\}$. That is, since $X_1, \cdots, X_n$ is another basis for the space $span\{\epsilon_1, \cdots, \epsilon_n\}$, we can write

$$\epsilon_n = b_0X_n + b_1X_{n-1} + b_2X_{n-2} + \cdots + b_{n-1}X_1, \tag{4.2.2}$$

with $b_0 = 1$, $b_i \in \mathbb{R}$.

Then by taking the conditional expectation of (4.2.2) given $\mathcal{F}_{n-1}$,

$$0 = E[\epsilon_n|\mathcal{F}_{n-1}] = E[b_0X_n + b_1X_{n-1} + b_2X_{n-2} + \cdots + b_{n-1}X_1|\mathcal{F}_{n-1}].$$

Since $b_0 = 1$ and $X_{n-1}, \cdots, X_1$ are all $\mathcal{F}_{n-1}$-measurable, we have

$$E[X_n|\mathcal{F}_{n-1}] = -\left(b_1X_{n-1} + b_2X_{n-2} + \cdots + b_{n-1}X_1\right). \tag{4.2.3}$$

We will use this below to construct processes that almost satisfy (2.0.3) without the need for much computation.

We will need the following result about power series; see [5] for a proof.

Lemma 4.2.1. *Let* $f(x) = \sum_{n\geq 0} a_n x^n$ *and* $g(x) = \sum_{n\geq 0} b_n x^n$ *be two power series. Then* $f(x)g(x) \equiv 1$ *if and only if*

$$\frac{b_n}{b_0} = (-a_0)^{-n} det \begin{pmatrix} a_1 & a_0 & & \\ a_2 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & a_0 \\ a_n & \cdots & a_2 & a_1 \end{pmatrix}.$$

We employ this result to prove the following.

Theorem 4.2.2. *Given sequences of numbers* $\{v_1, v_2, \dots\}$ *and* $\{a_0, a_1, a_2, \cdots\}$ *with* $a_0 \neq 0$, *define* $u_n = a_0 v_n + a_1 v_{n-1} + a_2 v_{n-2} + \cdots + a_{n-1} v_1$ *for* $n \geq 1$, *and let*

$$A_n = \begin{bmatrix} a_0 & & & \\ a_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ a_{n-1} & \cdots & a_1 & a_0 \end{bmatrix}.$$

Then A_n^{-1} *has the form*

$$A_n^{-1} = \begin{bmatrix} b_0 & & & \\ b_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ b_{n-1} & \cdots & b_1 & b_0 \end{bmatrix},$$

and $v_n = b_0 u_n + b_1 u_{n-1} + b_2 u_{n-2} + \cdots + b_{n-1} u_1$ *for each* $n \geq 1$*, where the coefficients* $\{b_n\}$ *satisfy* $\left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) \equiv 1$*.*

Proof. We have

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} = A_n \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

for each $n \geq 1$. Since $A_n^{-1} = \frac{1}{det(A_n)} adj(A_n)$ where $adj(A_n)$ is the adjugate matrix of A_n, it is easy to see A_n^{-1} has the stated form and thus the given representation for v_n holds. It remains to show

$$\left(\sum_{n=0}^{\infty} a_n x^n\right)\left(\sum_{n=0}^{\infty} b_n x^n\right) \equiv 1.$$

For the $(n,1)$ entry of A_n^{-1},

$$b_{n-1} = \frac{(-1)^{n+1}}{det(A)} det \begin{pmatrix} a_1 & a_0 & & \\ a_2 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & a_0 \\ a_{n-1} & \cdots & a_2 & a_1 \end{pmatrix} = (-a_0)^{1-n} det \begin{pmatrix} a_1 & a_0 & & \\ a_2 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & a_0 \\ a_{n-1} & \cdots & a_2 & a_1 \end{pmatrix}.$$

Apply the previous lemma, we have the result. □

We can apply the Theorem 4.2.2 to construct processes $\{X_n\}$ that behave almost like a GSAP. If $\{X_n\}$ are defined as in (4.2.1), then they also satisfy (4.2.3). We now give some examples of this construction.

Example 4.2.1. *Suppose we want to build a process $\{X_n\}$ such that*

$$E[X_n|\mathcal{F}_{n-1}] = \frac{1}{2}X_{n-1} + \frac{1}{2}X_{n-2}.$$

Then $b_1 = b_2 = -\frac{1}{2}$ and $b_i = 0$ for $i \geq 3$. So the corresponding power series becomes

$$\sum_{n\geq 0} b_n x^n = 1 - \frac{1}{2}x - \frac{1}{2}x^2.$$

Then the complementary power series is

$$\sum_{n\geq 0} a_n x^n = \frac{1}{1 - \frac{1}{2}x - \frac{1}{2}x^2} = \sum_{n=0}^{\infty} \frac{1}{3}\left(2 + (-\frac{1}{2})^n\right) x^n = 1 + \frac{1}{2}x + \frac{3}{4}x^2 + \frac{5}{8}x^3 + \frac{11}{16}x^4 + \cdots$$

So the sequence $\{a_n\}_{n\geq 0}$ is given by $a_n = \frac{1}{3}\left(2 + (-\frac{1}{2})^n\right)$. Thus if we let $X_n = \sum_{i=0}^{n-1} \frac{1}{3}\left(2 + (-\frac{1}{2})^i\right) \epsilon_{n-i}$, where $\{\epsilon_i\}$ are i.i.d mean-zero random variables, we have a process $\{X_n\}$ such that for $n \geq 3$,

$$E[X_n|\mathcal{F}_{n-1}] = \frac{1}{2}X_{n-1} + \frac{1}{2}X_{n-2}.$$

Of course, this is not technically a GSAP because for $n = 2$,

$$E[X_2|\mathcal{F}_1] = \frac{1}{2}X_1,$$

which does not satisfy (2.0.3). *But* (2.0.3) *is satisfied for all $n > 2$ and so the process $\{X_n\}$ eventually behaves like a GSAP.*

Thus one can easily construct processes $\{X_n\}$ that satisfy (2.0.3) except for the first several terms.

We now give another example of construction (4.2.1).

Example 4.2.2. *Suppose we accumulate some input from time $t = 0$ on. ϵ_i is the input at $t = i$ and its influence will decrease with time. Let's assume the ϵ_i are i.i.d mean 0 and their influence is inversely proportional to the time difference, and X_i is the cumulative effect at time $t = i$. Then*

$$X_i = \epsilon_i + \frac{1}{2}\epsilon_{i-1} + \cdots + \frac{1}{i}\epsilon_1. \tag{4.2.4}$$

Then compared with (4.2.1), we can see $a_i = \frac{1}{i+1}$ so the power series is

$$\sum_{n\geq 0} a_n x^n = \sum_{n\geq 0} \frac{1}{n+1} x^n = \frac{-\ln(1-x)}{x}.$$

By the previous theorem, the power series

$$\sum_{n\geq 0} b_n x^n = \frac{-x}{\ln(1-x)}.$$

We will later show

$$b_1 < b_2 < b_3 < \cdots < b_n < \cdots < 0 \tag{4.2.5}$$

and

$$\sum_{n=1}^{\infty} b_n = -1 \tag{4.2.6}$$

So by (4.2.3),

$$E[X_n|\mathcal{F}_{n-1}] = -b_1 X_{n-1} - b_2 X_{n-2} - \cdots - b_{n-1} X_1.$$

We will show the conditional expectation is not a weighted average but it is in a limiting sense.

We first prove a lemma for later use.

Lemma 4.2.3. *If* $f(x) = \frac{-\ln(1-x)}{x}$. *Then we can write*

$$\frac{1}{f(x)} = \frac{x}{-\ln(1-x)} = \int_0^1 (1-x)^t dt$$

Proof. When $0 < x < 1$,

$$\int_0^1 (1-x)^t dt = \int_0^1 e^{t\ln(1-x)} dt = \frac{1}{\ln(1-x)} e^{t\ln(1-x)}\Big|_0^1 = \frac{x}{-\ln(1-x)}$$

We can see when $x \to 0$ or $x \to 1$, the result still holds. □

Now we can show the following.

Lemma 4.2.4. $b_1 < b_2 < b_3 < \cdots < b_n < \cdots < 0$ *and* $\sum_{n=1}^{\infty} b_n = -1$ *for* b_n *given in*

$$h(x) = \sum_{n\geq 0} b_n x^n = \frac{-x}{\ln(1-x)}$$

Proof. For the negativity of $\{b_n\}$, it suffices to show every order derivative of the function $h(x)$ is negative. From the previous lemma,

$$h(x) = \frac{-x}{\ln(1-x)} = \int_0^1 (1-x)^t dt.$$

Differentiating under integral we have

$$h'(x) = \int_0^1 (-t)(1-x)^{t-1} dt,$$

hence $h'(0) < 0$.

Differentiate again we have

$$h''(x) = \int_0^1 (-t)(1-t)(1-x)^{t-1} dt,$$

hence $h''(0) < 0$.

Repeating the process we can show every derivative of h at $x = 0$ is negative.

Note the monotonicity of b_n is equivalent to every higher-than-1 order derivative of the function $(1-x)h(x)$ is positive.

Note that around $x = 0$,

$$(1-x)h(x) = \frac{-x(1-x)}{\ln(1-x)} = \int_1^2 (1-x)^t dt$$

which follows by similar arguments. Also note $h(1) = 0$ hence $\sum_{n=1}^{\infty} b_n = -1$. □

So (4.2.4) constructs a process $\{X_n\}$ such that the conditional expectation is a linear combination of the entire history with more recent terms getting higher weights. And the linear combination is a weighted average in limit sense.

Chapter 5

Maximal Inequalities

5.1 GSAPs that are bounded below

Analogous to Theorem 1.0.2 and Theorem 1.0.3, we can show a maximal inequality for GSAPs. Throughout this section, we assume X_n is a non-negative GSAP. In general if $\{X_n\}$ is bounded below by l then we could consider the GSAP $X_n - l$.

Recall that $X_n^* = \max_{1 \le k \le n} X_k$.

Theorem 5.1.1. *(Maximal Inequality) If $\{X_i\}$ is a non-negative GSAP, then for all $x > 0$,*

$$xP(X_n^* \ge x) \le M_n E[X_n; X_n^* \ge x] \le M_n E[X_n] \tag{5.1.1}$$

where M_n is a constant depending only on the triangular array T up to the $(n-1)^{st}$ row.

Proof. Let $A_k = \{X_{k-1}^* < x \le X_k\} \in \mathcal{F}_k$ be the event that the process exceeds level x at time k for the first time. Then $\{X_n^* \ge x\}$ is the disjoint union of all the A_k's for

$k = 1, 2, \cdots, n$. Hence

$$xP(X_n^* \geq x) = x \sum_{k=1}^{n} P(A_k) \leq \sum_{k=1}^{n} E[X_k; A_k].$$

Since

$$\tau_k^{n,k} X_k \leq E[X_n | \mathcal{F}_k] = \tau_k^{n,k} X_k + \cdots + \tau_1^{n,k} X_1 \tag{5.1.2}$$

on A_k,

$$\sum_{k=1}^{n} E[X_k; A_k] \leq \sum_{k=1}^{n} E[\frac{E[X_n|\mathcal{F}_k]}{\tau_k^{n,k}}; A_k] = \sum_{k=1}^{n} \frac{1}{\tau_k^{n,k}} E[X_n; A_k].$$

Let $M_n = \max\{\frac{1}{\tau_k^{n,k}} : k = 1, 2, \cdots, n\}$, and note that $\tau_k^{n,k} \neq 0$ for all $k = 1, 2, \cdots, n$ since $\tau_k^{n,k} \geq \tau_{k-1}^{n,k} \geq \cdots \geq \tau_1^{n,k}$ and their sum is one. Then

$$\sum_{k=1}^{n} \frac{1}{\tau_k^{n,k}} E[X_n; A_k] \leq \sum_{k=1}^{n} M_n E[X_n; A_k] = M_n E[X_n; X_n^* \geq x] \leq M_n E[X_n].$$

□

Theorem 5.1.2. *(Other Forms of the Maximal Inequality)*

Suppose $\{X_i\}$ is a non-negative GSAP and let $p > 0$. Then for all $x > 0$,

$$x^p P(X_n^* \geq x) \leq M_n^p E[X_n^p],$$

where M_n is the same constant as above.

Proof. Take $A_k = \{X_{k-1}^* < x \leq X_k\} \in \mathcal{F}_k$ as before. Then letting $Y_k = E[X_n|\mathcal{F}_k]$,

$$\begin{aligned} x^p P(X_n^* \geq x) &= x^p \sum_{k=1}^{n} P(A_k) \leq \sum_{k=1}^{n} E[X_k^p; A_k] \leq \sum_{k=1}^{n} E[M_n^p Y_k^p; A_k] \\ &= \sum_{k=1}^{n} M_n^p E[Y_k^p; A_k] \leq M_n^p E[Y_n^p] = M_n^p E[X_n^p] \end{aligned}$$

□

Since the constant M_n depends on n, it would be useful to have condition under which $\sup_n M_n$ is finite.

Theorem 5.1.3. *If* $\frac{\tau_{i-1}^{n,n-1}}{\tau_i^{n,n-1}} < q < 1$ *for all* n *and* $i < n$ *(that is if every row of the triangular array* T *decreases exponentially fast) then* $M = \sup_n M_n < \infty$*.*

Proof. We first show by induction that $\frac{\tau_{j-1}^{n,k}}{\tau_j^{n,k}} \le q$, for all $n > k \ge j > 1$. In other words $\tau_k^{n,k}, \tau_{k-1}^{n,k}, \cdots, \tau_1^{n,k}$ is still decreasing exponentially fast. Fix $n > 1$, clearly it is true when $k = n-1$ by assumption. Suppose $\tau_k^{n,k}, \tau_{k-1}^{n,k}, \cdots, \tau_1^{n,k}$ is decreasing exponentially fast. Since $\tau_i^{n,k-1} = \tau_i^{n,k} + \tau_k^{n,k}\tau_i^{k,k-1}$, the relative ratio

$$\frac{\tau_{i-1}^{n,k-1}}{\tau_i^{n,k-1}} = \frac{\tau_{i-1}^{n,k} + \tau_k^{n,k}\tau_{i-1}^{k,k-1}}{\tau_i^{n,k} + \tau_k^{n,k}\tau_i^{k,k-1}} = \frac{\tau_i^{n,k}\frac{\tau_{i-1}^{n,k}}{\tau_i^{n,k}} + \tau_k^{n,k}\tau_i^{k,k-1}\frac{\tau_{i-1}^{k,k-1}}{\tau_i^{k,k-1}}}{\tau_i^{n,k} + \tau_k^{n,k}\tau_i^{k,k-1}}$$

is a weighted average of $\frac{\tau_{i-1}^{n,k}}{\tau_i^{n,k}}$ and $\frac{\tau_{i-1}^{k,k-1}}{\tau_i^{k,k-1}}$, both of which are $\le q$ so $\frac{\tau_{i-1}^{n,k-1}}{\tau_i^{n,k-1}} \le q$.

Moreover observe that for all $n > 1$,

$$1 = \tau_k^{n,k} + \cdots + \tau_1^{n,k} \le \tau_k^{n,k} + q\tau_k^{n,k} + q^2\tau_k^{n,k} + \cdots + q^{k-1}\tau_k^{n,k} < \frac{1}{1-q}\tau_k^{n,k}.$$

Hence $\tau_k^{n,k} > 1-q$, and $M_n = \max_k\{\frac{1}{\tau_k^{n,k}}\} < \frac{1}{1-q}$. Thus $M = \sup_n M_n \le \frac{1}{1-q}$. □

Corollary 5.1.4. *If the array* T *in given by* $\{c_n\}$ *as in* (2.0.2)*, and if for all* n*,*

$$\frac{c_{n+1}}{c_n} \le q < 1,$$

then the condition of the previous theorem holds directly from (2.0.2)*, and hence* M_n *has a uniform bound.*

For the next result, we need the following dimple generalization of a standard formula for moments of positive random variables.

Lemma 5.1.5. *Let W, Z be positive random variables. Then for all $r > 0$,*

$$E[WZ^r] = r\int_0^\infty x^{r-1}E[W; Z > x]dx$$

Proof. Letting F_{zw} denote the joint law of W and Z,

$$\begin{aligned} E[WZ^r] &= \int_{[0,\infty]\times[0,\infty]} wz^r dF_{zw} \\ &= \int_{[0,\infty]\times[0,\infty]} w[\int_0^z rx^{r-1}dx]dF_{zw} \\ &= \int_0^\infty rx^{r-1}[\int_{[x,\infty]\times[0,\infty]} wdF_{zw}]dx \\ &= \int_0^\infty rx^{r-1}E[W; Z > x]dx, \end{aligned}$$

where we changed the order of integration by Tonelli's Theorem since all terms are positive. □

We now prove on L^p version of the maximal inequality for GSAP. Recall that $X_n \geq 0$ for all n.

Theorem 5.1.6. *(L^p Maximal Inequality)*

For $p > 1$, $||X_n^*||_p \leq \frac{p}{p-1}M_n||X_n||_p$.

Proof. By the previous lemma using $W = 1$, the maximal inequality (5.1.1), and Holder's inequality,

$$
\begin{aligned}
\|X_n^*\|_p^p = E[(X_n^*)^p] &= p\int_0^\infty t^{p-1}P(X_n^* > t)dt \\
&\leq p\int_0^\infty t^{p-1}\frac{M_n E[X_n; X_n^* > t]}{t}dt \\
&= pM_n\int_0^\infty t^{p-2}E[X_n; X_n^* > t]dt \\
&= \frac{pM_n}{p-1}E[X_n(X_n^*)^{p-1}] \leq \frac{pM_n}{p-1}\|X_n\|_p\|(X_n^*)^{p-1}\|_q \\
&= \frac{pM_n}{p-1}\|X_n\|_p(E[(X_n^*)^{(p-1)q}])^{1/q} = \frac{pM_n}{p-1}\|X_n\|_p(E[(X_n^*)^p])^{1/q} \\
&= \frac{pM_n}{p-1}\|X_n\|_p\|X_n^*\|_p^{p/q},
\end{aligned}
$$

where $\frac{1}{p} + \frac{1}{q} = 1$. So considering the leftmost and rightmost expression, we have

$$\|X_n^*\|_p = \|X_n^*\|_p^{p-p/q} \leq \frac{pM_n}{p-1}\|X_n\|_p.$$

□

The proof of the following is directly analogous to the proof of Theorem 5.1.1.

Theorem 5.1.7. *(Exponential Bound) If $\{X_n\}$ is a non-negative GSAP, then*

$$P(X_n^* \geq a) \leq e^{-\lambda a}E[e^{\lambda M_n X_n}]$$

for all $n > 1$, $\lambda > 0$ and all $a > 0$.

Proof.

$$P(X_n^* \geq a) = P(e^{\lambda X_n^*} \geq e^{\lambda a}) = \sum_{k=1}^{n} P(A_k)$$

where $A_k \in \mathcal{F}_k$ is the event that the process $e^{\lambda X_n}$ first exceeds $e^{\lambda a}$ at time k. Then

$$\sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} E[1; A_k] \leq \sum_{k=1}^{n} E[\frac{e^{\lambda X_k}}{e^{\lambda a}}; A_k] = e^{-\lambda a} \sum_{k=1}^{n} E[e^{\lambda X_k}; A_k]$$

Let $Y_k = E[X_n|\mathcal{F}_k]$ then Y_k is a martingale and $Y_n = X_n$.

From (5.1.2) we can see $X_k \leq M_n Y_k$ on A_k. So

$$e^{-\lambda a} \sum_{k=1}^{n} E[e^{\lambda X_k}; A_k] \leq e^{-\lambda a} \sum_{k=1}^{n} E[e^{\lambda M_n Y_k}; A_k].$$

Since $y \mapsto e^{\lambda M_n y}$ is convex, $e^{\lambda M_n Y_k}$ is a submartingale. So

$$\begin{aligned} e^{-\lambda a} \sum_{k=1}^{n} E[e^{\lambda M_n Y_k}; A_k] &\leq e^{-\lambda a} \sum_{k=1}^{n} E[e^{\lambda M_n Y_n}; A_k] \\ &\leq e^{-\lambda a} E[e^{\lambda M_n Y_n}] = e^{-\lambda a} E[e^{\lambda M_n X_n}] \end{aligned}$$

□

5.2 GSAPs with bounded increments

If the GSAP $\{X_n\}$ has bounded increments, we can establish similar results.

Theorem 5.2.1. *(Maximal Inequality) Suppose $\{X_n\}$ is a GSAP with bounded increments, namely $|X_n - X_{n-1}| \leq \delta$ for all n, then for all $x > 0$,*

$$xP(X_n^* \geq x) \leq E[X_n] + N_n \delta \tag{5.2.1}$$

where $N_n \leq \frac{n}{2}$ depends on T up to the $(n-1)^{st}$ row.

Proof. Again let $A_k = \{X^*_{k-1} < x \le X_k\} \in \mathcal{F}_k$ be the event that the process exceeds level x at time k for the first time. Then $\{X^*_n \ge x\}$ is the disjoint union of all the A_k for $k = 1, 2, \cdots, n$. Hence

$$xP(X^*_n \ge x) = x\sum_{k=1}^{n} P(A_k) \le \sum_{k=1}^{n} E[X_k; A_k].$$

Since

$$\begin{aligned} &\tau^{n,k}_k X_k + \tau^{n,k}_{k-1}(X_k - \delta) + \cdots + \tau^{n,k}_1 (X_k - (k-1)\delta) \\ &\le \tau^{n,k}_k X_k + \tau^{n,k}_{k-1} X_{k-1} + \cdots + \tau^{n,k}_1 X_1 = E[X_n|\mathcal{F}_k], \end{aligned}$$

we have on A_k

$$X_k - \left(\tau^{n,k}_{k-1} + \cdots + (k-1)\tau^{n,k}_1\right)\delta \le E[X_n|\mathcal{F}_k]$$

on A_k. Let

$$N_n = \max\{\tau^{n,k}_{k-1} + \cdots + (k-1)\tau^{n,k}_1 : k = 1, 2, \cdots, n\}.$$

By Lemma 3.0.1, $\tau^{n,k}_i$ are non-increasing and sum to one. So

$$\tau^{n,k}_{k-1} + \cdots + (k-1)\tau^{n,k}_1 \le \frac{k}{2},$$

and hence $N_n \le \frac{n}{2}$. Then

$$\begin{aligned} \sum_{k=1}^{n} E[X_k; A_k] &\le \sum_{k=1}^{n} E[E[X_n|\mathcal{F}_k] + \left(\tau^{n,k}_{k-1} + \cdots + (k-1)\tau^{n,k}_1\right)\delta; A_k] \\ &\le \sum_{k=1}^{n} E[E[X_n|\mathcal{F}_k] + N_n\delta; A_k] \le E[X_n] + N_n\delta. \end{aligned}$$

□

Note that in the previous proof we use the rough bound

$$\tau_{k-1}^{n,k} + \cdots + (k-1)\tau_1^{n,k} \le \frac{k}{2}.$$

So if the τ's decrease exponentially fast, then we can get a uniform bound for N_n.

Theorem 5.2.2. *(Ratio Test) If* $\frac{\tau_{i-1}^{n,n-1}}{\tau_i^{n,n-1}} < q < 1$ *for all* n *and all* $i < n$ *(namely if every row of the triangular array* T *is decreasing exponentially fast) then* $N = \sup_n N_n < \infty$.

Proof. As in the proof of Theorem 5.1.3, we know $\tau_k^{n,k}, \tau_{k-1}^{n,k}, \cdots, \tau_1^{n,k}$ is decreasing exponentially fast. Then

$$\tau_{k-1}^{n,k} + \cdots + (k-1)\tau_1^{n,k} \le \tau_{k-1}^{n,k} + 2q\tau_{k-1}^{n,k} + \cdots + (k-2)q^{k-1}\tau_{k-1}^{n,k}$$

$$= \tau_{k-1}^{n,k}[1 + 2q + \cdots + (k-2)q^{k-1}] \le \tau_{k-1}^{n,k}\frac{1-q^k}{(1-q)^2}$$

Also we know if $\tau_k^{n,k}, \tau_{k-1}^{n,k}, \cdots, \tau_1^{n,k}$ is decreasing exponentially fast then the second term $\tau_{k-1}^{n,k} \le q$ and hence

$$\tau_{k-1}^{n,k} + \cdots + (k-1)\tau_1^{n,k} \le q\frac{1-q^k}{(1-q)^2}.$$

Then we have

$$N_n = \max\{\tau_{k-1}^{n,k} + \cdots + (k-1)\tau_1^{n,k} : k = 1, 2, \cdots, n\} \le \frac{q}{(1-q)^2}.$$

So $N = \sup_n N_n \le \frac{q}{(1-q)^2} < \infty$ □

Corollary 5.2.3. *Again if the array* T *is given by* $\{c_n\}$ *as in* (2.0.2)*, and for all* n*,*

$$\frac{c_{n+1}}{c_n} \le q < 1,$$

then the condition of the previous theorem holds directly from (2.0.2), *and hence* N_n *has a uniform bound.*

For classic martingales with bounded increments, the following theorem is well known.

Theorem 5.2.4. *(Azuma Inequality) Suppose* $\{X_k\}$ *is a martingale with* $|X_k - X_{k-1}| \leq a_k$ *for each* k*. Then*

$$P(|X_n - X_0| \geq t) \leq 2\exp(\frac{-t^2}{2\sum_1^n a_k^2}).$$

We can prove a similar result for GSAPs.

Theorem 5.2.5. *(Azuma Inequality for GSAPs) Suppose* $\{X_k\}$ *is a GSAP with* $|X_k - X_{k-1}| \leq a_k$ *for each* k*. Then*

$$P(|X_n - X_0| \geq t) \leq 2\exp(\frac{-t^2}{2\sum_1^n b_k^2}),$$

where

$$b_k = \tau_k^{n,k}[a_k + (1 - \tau_{k-1}^{k,k-1})a_{k-1} + \cdots + (1 - \tau_{k-1}^{k,k-1} - \cdots - \tau_2^{k,k-1})a_2].$$

Proof. Take $Y_k = E[X_n|\mathcal{F}_k]$ so that Y_k is a martingale. In particular $Y_0 = X_0$, $Y_1 = X_1$,

$Y_n = X_n$. For $k \geq 2$,

$$
\begin{aligned}
|Y_k - Y_{k-1}| &= |E[X_n|\mathcal{F}_k] - E[X_n|\mathcal{F}_k|\mathcal{F}_{k-1}]| \\
&= |\tau_k^{n,k} X_k + \cdots + \tau_1^{n,k} X_1 - (\tau_{k-1}^{n,k} + \tau_k^{n,k}\tau_{k-1}^{k,k-1})X_{k-1} - \cdots - (\tau_1^{n,k} + \tau_k^{n,k}\tau_1^{k,k-1})X_1| \\
&= |\tau_k^{n,k} X_k - \tau_k^{n,k}\tau_{k-1}^{k,k-1} X_{k-1} - \cdots - \tau_k^{n,k}\tau_1^{k,k-1} X_1| \\
&= \tau_k^{n,k} |X_k - \tau_{k-1}^{k,k-1} X_{k-1} - \cdots - \tau_1^{k,k-1} X_1| \\
&= \tau_k^{n,k} |(X_k - X_{k-1}) + (1 - \tau_{k-1}^{k,k-1})(X_{k-1} - X_{k-2}) \\
&\quad + \cdots + (1 - \tau_{k-1}^{k,k-1} - \cdots - \tau_2^{k,k-1})(X_2 - X_1)| \\
&\leq \tau_k^{n,k}[a_k + (1 - \tau_{k-1}^{k,k-1})a_{k-1} + \cdots + (1 - \tau_{k-1}^{k,k-1} - \cdots - \tau_2^{k,k-1})a_2] = b_k,
\end{aligned}
$$

where the last equality serves as the definition of b_k. Define $b_1 = a_1$ then apply the classic Azuma inequality to the martingale $\{Y_k\}$. □

Chapter 6

Convergence of GSAPs

For convenience, we still write

$$E[X_n|\mathcal{F}_m] = \tau_m^{n,m} X_m + \cdots + \tau_1^{n,m} X_1$$

for $n < m$, in which case only $\tau_n^{n,m} = 1$, and for $i \neq n$, $\tau_i^{n,m} = 0$. (Note that in this case the non-increasing property is no longer true.)

If $\{X_n\}$ is a GSAP with respect to T, then by tower property, we have for $n \geq k$:

$$\begin{aligned} E[X_{n+1}|\mathcal{F}_k] &= E[X_{n+1}|\mathcal{F}_n|\mathcal{F}_k] \\ &= E[\tau_n^{n+1,n} X_n + \cdots + \tau_1^{n+1,n} X_1|\mathcal{F}_k] \\ &= \tau_n^{n+1,n} E[X_n|\mathcal{F}_k] + \cdots + \tau_1^{n+1,n} E[X_1|\mathcal{F}_k] \\ &= \tau_n^{n+1,n}(\tau_k^{n,k} X_k + \cdots + \tau_1^{n,k} X_1) + \cdots + \tau_1^{n+1,n}(\tau_k^{1,k} X_k + \cdots + \tau_1^{1,k} X_1) \end{aligned}$$

and as in Lemma 3.0.1,

$$E[X_{n+1}|\mathcal{F}_k] = \tau_k^{n+1,k} X_k + \cdots + \tau_1^{n+1,k} X_1.$$

By comparing terms we can see for $i \le k$,

$$\tau_i^{n+1,k} = \tau_n^{n+1,n}\tau_i^{n,k} + \cdots + \tau_1^{n+1,n}\tau_i^{1,k}. \tag{6.0.1}$$

So for fixed k, i, the sequence of numbers $\{\tau_i^{n,k}\}_{n\ge1}$ has the property that the next term is a weighted average of the previous terms.

In general consider a sequence of numbers $\{a_n\}$ with

$$a_{n+1} = \tau_n^{n+1,n}a_n + \cdots + \tau_1^{n+1,n}a_1 \tag{6.0.2}$$

We will first show the sequence converges. We begin with a result needed later.

Lemma 6.0.1. *For* $1 \ge u, v \ge 0$, *and all* $a, b, c \in \mathbb{R}$

$$|ua + (1-u)b - va - (1-v)c| \le \max\{1-v, 1-u\} \cdot \max\{|b-c|, |a-c|, |a-b|\}$$

Proof. If $u \ge v$,

$$\begin{aligned} |ua + (1-u)b - va - (1-v)c| &= |(u-v)a + (1-u)b - (1-v)c| \\ &= |(1-u)(b-c) + (u-v)(c-a)| \\ &\le (1-u)|b-c| + (u-v)|c-a| \\ &\le (1-v)\max\{|b-c|, |a-c|\} \\ &\le \max\{1-v, 1-u\} \cdot \max\{|b-c|, |a-c|, |a-b|\} \end{aligned}$$

When $u < v$, we can use the same argument. □

Now we prove the sequence in (6.0.2) converges.

Theorem 6.0.2. *(Convergence of Weighted Average Sequences) Let T be a self-averaging array and set $a_1, a_2, \cdots, a_k$ as initial conditions. Then for $n \geq k$, if*

$$a_{n+1} = \tau_n^{n+1,n} a_n + \cdots + \tau_1^{n+1,n} a_1,$$

then the sequence $\{a_n\}$ converges.

Proof. Let A_n be the set of monotone weighted averages of $\{a_1, \cdots a_n\}$:

$$A_n = \{w_1 a_n + w_2 a_{n-1} + \cdots w_n a_1\},$$

where $w_1 \geq w_2 \geq \cdots \geq w_n$ and $w_1 + w_2 + \cdots + w_n = 1$. Let

$$D_n = \sup\{|x - y| : x, y \in A_n\}.$$

We can see A_n is a convex set and $a_{n+1} \in A_n$ by definition. Also

$$a_{n+2} = \tau_{n+1}^{n+2,n+1} a_{n+1} + \cdots + \tau_1^{n+2,n+1} a_1$$

is a weighted average of a_{n+1} and $\frac{\tau_n^{n+2,n+1}}{\tau_n^{n+2,n+1} + \cdots + \tau_1^{n+2,n+1}} a_n + \cdots + \frac{\tau_1^{n+2,n+1}}{\tau_n^{n+2,n+1} + \cdots + \tau_1^{n+2,n+1}} a_1$, so $a_{n+2} \in A_n$. Recursively we can show $a_{n+i} \in A_n$ for $i = 1, 2, 3, \cdots$.

Take any two elements in A_n $(n \geq k)$: $w_1 a_n + w_2 a_{n-1} + \cdots w_n a_1$ and $w'_1 a_n + w'_2 a_{n-1} + \cdots w'_n a_1$. Then by Lemma 6.0.1, using $u = w_1$ and $v = w'_1$,

$$\begin{aligned} &|w_1 a_n + w_2 a_{n-1} + \cdots w_n a_1 - (w'_1 a_n + w'_2 a_{n-1} + \cdots w'_n a_1)| \\ \leq \max\{1 - w_1, 1 - w'_1\} \cdot \max\{&|a_n - \frac{w_2}{w_2 + \cdots + w_n} a_{n-1} - \cdots - \frac{w_n}{w_2 + \cdots + w_n} a_1|, \\ &|a_n - \frac{w'_2}{w'_2 + \cdots + w'_n} a_{n-1} - \cdots - \frac{w'_n}{w'_2 + \cdots + w'_n} a_1|, \\ |\frac{w_2}{w_2 + \cdots + w_n} a_{n-1} + \cdots + \frac{w_n}{w_2 + \cdots + w_n} a_1 &- \frac{w'_2}{w'_2 + \cdots + w'_n} a_{n-1} - \cdots - \frac{w'_n}{w'_2 + \cdots + w'_n} a_1|\} \end{aligned}$$

Note that w_1 and w_1' are the largest among $\{w_1, \cdots w_n\}$ and $\{w_1', \cdots w_n'\}$ respectively so

$$\max\{1 - w_1, 1 - w_1'\} \le \frac{n-1}{n}.$$

Also note that all of the three terms

$$a_n, \frac{w_2}{w_2 + \cdots + w_n} a_{n-1} + \cdots + \frac{w_n}{w_2 + \cdots + w_n} a_1, \frac{w_2'}{w_2' + \cdots + w_n'} a_{n-1} + \cdots + \frac{w_n'}{w_2' + \cdots + w_n'} a_1$$

are elements of A_{n-1}. So $|w_1 a_n + w_2 a_{n-1} + \cdots + w_n a_1 - (w_1' a_n + w_2' a_{n-1} + \cdots + w_n' a_1)| \le \frac{n-1}{n} D_{n-1}$ and hence $D_n \le \frac{n-1}{n} D_{n-1}$. So we have $D_n \to 0$.

For $m > n$ large enough, a_m and a_n are both in A_n so

$|a_m - a_n| \le D_n \to 0$ so the sequence is Cauchy thus convergent. □

Armed with this theorem we can prove a result for the conditional expectations of a GSAP.

Theorem 6.0.3. *(Convergence of conditional expectations)*

Fix k and $i \le k$. Then $\{\tau_i^{n,k}\}_{n\ge0}$ in (6.0.1) is convergent by Theorem 6.0.2. Denote the limit by $\tau_i^{\infty,k}$. Let $\{X_n\}$ be a GSAP for the array T. Then for each k, there exist constants $\tau_1^{\infty,k}, \cdots, \tau_k^{\infty,k}$ such that

$$\lim_{n\to\infty} E[X_n|\mathcal{F}_k] = \tau_k^{\infty,k} X_k + \cdots + \tau_1^{\infty,k} X_1 \quad a.s.$$

Proof. Since $\tau_i^{n,k} \to \tau_i^{\infty,k}$, then

$$E[X_n|\mathcal{F}_k] = \tau_k^{n,k} X_k + \cdots + \tau_1^{n,k} X_1 \to \tau_k^{\infty,k} X_k + \cdots + \tau_1^{\infty,k} X_1 \quad a.s.$$

□

Now we consider convergence of a GSAP $\{X_n\}$.

Take $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$. Recall in Theorem 1.0.5, a classical martingale $\{X_n\}$ is bounded in L^2 if and only if $\sum E[(X_n - X_{n-1})^2] = \sum E|\epsilon_n|^2 < \infty$. We can prove an analogous theorem for GSAPs.

Observe that for $n > m$,

$$E|X_n - X_m|^2 = E|X_n - E[X_n|\mathcal{F}_m]|^2 + E|E[X_n|\mathcal{F}_m] - X_m|^2. \tag{6.0.3}$$

So for a GSAP to converge in L^2, it suffices to show both terms on the right hand side converge to zero. We show the first term converges to zero in the following theorem.

Theorem 6.0.4. *Let $\{X_n\}$ be a GSAP for the array T. If $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$ satisfies $\sum E|\epsilon_n|^2 < \infty$, then $E|X_n - E[X_n|\mathcal{F}_m]|^2 \to 0$ for $n \geq m \to \infty$.*

Proof. Since $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$, it is $\mathcal{F}_n$-measurable and $E[\epsilon_n|\mathcal{F}_{n-1}] = 0$. By Theorem 4.1.1,

$$X_n = E[X_n|\mathcal{F}_m] + \epsilon_n + \tau_{n-1}^{n,n-1}\epsilon_{n-1} + \cdots + \tau_{m+1}^{n,m+1}\epsilon_{m+1}.$$

We can see the terms on the right hand side are orthogonal to each other and hence

$$\begin{aligned} E|X_n - E[X_n|\mathcal{F}_m]|^2 &= E|\epsilon_n|^2 + (\tau_{n-1}^{n,n-1})^2 E|\epsilon_{n-1}|^2 + \cdots + (\tau_{m+1}^{n,m+1})^2 E|\epsilon_{m+1}|^2 \\ &\leq E|\epsilon_n|^2 + E|\epsilon_{n-1}|^2 + \cdots + E|\epsilon_{m+1}|^2 \to 0 \end{aligned}$$

as $n \geq m \to \infty$. □

From the proof above we can see that a necessary condition for $\{X_n\}$ to converge in L^2 is $\sum \frac{1}{n^2} E|\epsilon_n|^2 < \infty$. Indeed, the right side of the first equality above is bounded below by

$$E|\epsilon_n|^2 + (\frac{1}{n-1})^2 E|\epsilon_{n-1}|^2 + \cdots + (\frac{1}{m+1})^2 E|\epsilon_{m+1}|^2.$$

In order to prove the second term in (6.0.3) converges to zero, we first prove the following.

Lemma 6.0.5. *If $a_n \to 0$ then $\frac{1}{n}(a_1 + \cdots + a_n) \to 0$.*

Proof. Fix $\epsilon > 0$. Then there exists N such that $|a_n| < \epsilon$ for $n > N$. Then

$$\begin{aligned}\frac{1}{n}|a_1 + \cdots + a_n| &= \frac{1}{n}|a_1 + \cdots + a_N + a_{N+1} + \cdots + a_n| \\ &\leq \frac{1}{n}|a_1 + \cdots + a_N| + \frac{1}{n}|a_{N+1} + \cdots + a_n|\end{aligned}$$

the first term $\frac{1}{n}|a_1 + \cdots + a_N| \to 0$ and the second term $\frac{1}{n}|a_{N+1} + \cdots + a_n| < \epsilon$, so

$$\frac{1}{n}(a_1 + \cdots + a_n) \to 0.$$

□

Now we can prove the L^2 convergence.

Theorem 6.0.6. *(L^2 Convergence)*

Let $\{X_n\}$ be a GSAP for the array T, and suppose $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$ satisfies $\sum E|\epsilon_n|^2 < \infty$. Then for $n \geq m \to \infty$, $E|X_m - E[X_n|\mathcal{F}_m]|^2 \to 0$ and hence $\{X_n\}$ converges in L^2

Proof. Define $A_n = \{w_1X_n + \cdots w_nX_1 : w_1 \geq w_2 \geq \cdots \geq w_n \text{ and } w_1 + w_2 + \cdots + w_n = 1\}$ and let $D_n = sup\{E|Y - Z|^2 : Y, Z \in A_n\}$. It is easy to see that $D_2 = E|\epsilon_2|^2$. Take any two elements in A_n: $w_1X_n + w_2X_{n-1} + \cdots + w_nX_1$ and $w'_1X_n + w'_2X_{n-1} + \cdots + w'_nX_1$. Without losing generality, suppose $w_1 \geq w'_1$, then

$$E|w_1X_n + w_2X_{n-1} + \cdots + w_nX_1 - (w'_1X_n + w'_2X_{n-1} + \cdots + w'_nX_1)|^2$$

$$= E|(w_1 - w'_1)X_n + w_2X_{n-1} + \cdots + w_nX_1 - (w'_2X_{n-1} + \cdots + w'_nX_1)|^2$$

$$= (1 - w'_1)^2 E|\frac{(w_1 - w'_1)X_n}{1 - w'_1} + \frac{w_2X_{n-1} + \cdots + w_nX_1}{1 - w'_1} - \frac{w'_2X_{n-1} + \cdots + w'_nX_1}{1 - w'_1}|^2$$

$$= (1 - w'_1)^2 E|\frac{(w_1 - w'_1)(E[X_n|\mathcal{F}_{n-1}] + \epsilon_n)}{1 - w'_1} + \frac{w_2X_{n-1} + \cdots + w_nX_1}{1 - w'_1} - \frac{w'_2X_{n-1} + \cdots + w'_nX_1}{1 - w'_1}|^2.$$

Note that ϵ_n satisfies $E[\epsilon_n|\mathcal{F}_{n-1}] = 0$ so

$$E|w_1X_n + w_2X_{n-1} + \cdots + w_nX_1 - (w'_1X_n + w'_2X_{n-1} + \cdots + w'_nX_1)|^2$$

$$= (1 - w'_1)^2 E|\frac{(w_1 - w'_1)(E[X_n|\mathcal{F}_{n-1}])}{1 - w'_1} + \frac{w_2X_{n-1} + \cdots + w_nX_1}{1 - w'_1} - \frac{w'_2X_{n-1} + \cdots + w'_nX_1}{1 - w'_1}|^2$$

$$+ (w_1 - w'_1)^2 E|\epsilon_n^2|.$$

Note the terms $\frac{(w_1 - w'_1)(E[X_n|\mathcal{F}_{n-1}])}{1 - w'_1} + \frac{w_2X_{n-1} + \cdots + w_nX_1}{1 - w'_1}$ and $\frac{w'_2X_{n-1} + \cdots + w'_nX_1}{1 - w'_1}$ are both in A_{n-1} so almost surely

$$D_n \leq (1 - w'_1)^2 D_{n-1} + (w_1 - w'_1)^2 E|\epsilon_n|^2 \leq (\frac{n-1}{n})^2 D_{n-1} + (\frac{n-1}{n})^2 E|\epsilon_n|^2.$$

Hence recursively we have a.s.

$$D_n \leq (\frac{n-1}{n})^2 E|\epsilon_n|^2 + \cdots + (\frac{2}{n})^2 E|\epsilon_3|^2 + (\frac{2}{n})^2 D_2$$
$$= (\frac{n-1}{n})^2 E|\epsilon_n|^2 + \cdots + (\frac{2}{n})^2 E|\epsilon_3|^2 + (\frac{2}{n})^2 E|\epsilon_2|^2$$

We will show $D_n \to 0$ a.s. as $n \to \infty$.

Let $\sigma_n = \sum_n^\infty E|\epsilon_n|^2$ then $\sigma_n \to 0$. We have

$$\begin{aligned}
(\frac{n-1}{n})^2 E|\epsilon_n|^2 + \cdots + (\frac{2}{n})^2 E|\epsilon_3|^2 + (\frac{2}{n})^2 E|\epsilon_2|^2 &= (\frac{1}{n})^2[\sum_2^n (k-1)^2 E|\epsilon_k|^2 + E|\epsilon_2|^2] \\
&\leq (\frac{1}{n})^2[\sum_2^n (2k-3)\sigma_k + E|\epsilon_2|^2] \\
&= \frac{1}{n}[\frac{2n-3}{n}\sigma_n + \cdots + \frac{1}{n}\sigma_2 + \frac{1}{n}E|\epsilon_2|^2] \\
&\leq \frac{1}{n}[2\sigma_n + \cdots + 2\sigma_2 + \frac{1}{n}E|\epsilon_2|^2] \to 0
\end{aligned}$$

by the previous lemma.

Thus, $D_n \to 0$ as $n \to \infty$.

Since X_m and $E[X_n|\mathcal{F}_m]$ are both in A_m,

$$E|X_m - E[X_n|\mathcal{F}_m]|^2 \leq D_m \to 0.$$

Hence $E|X_n - X_m|^2 = E|X_m - E[X_n|\mathcal{F}_m]|^2 + E|X_n - E[X_n|\mathcal{F}_m]|^2 \to 0.$

So X_n converges in L^2 □

For convergence in L^p, $p \neq 2$, we have the following theorem.

Theorem 6.0.7. *If* $\sum_{n=1}^\infty \|\epsilon_n\|_p < \infty$ *then* X_n *converges in* L^p. $p \geq 1$

Proof. Take $A_n = \{w_1 X_n + \cdots w_n X_1\}$ where $w_1 \geq w_2 \geq \cdots \geq w_n$ and $w_1 + w_2 + \cdots + w_n = 1$.

$D_n = sup\{\|Y - Z\|_p : Y, Z \in A_n\}$. Take any two elements in A_n,

$$\|w_1 X_n + w_2 X_{n-1} + \cdots + w_n X_1 - (w_1' X_n + w_2' X_{n-1} + \cdots + w_n' X_1)\|_p$$

$$= \|(w_1 - w_1')X_n + w_2 X_{n-1} + \cdots + w_n X_1 - (w_2' X_{n-1} + \cdots + w_n' X_1)\|_p$$

$$= (1 - w_1')\|\frac{(w_1 - w_1')(E[X_n|\mathcal{F}_{n-1}] + \epsilon_n)}{1 - w_1'} + \frac{w_2 X_{n-1} + \cdots + w_n X_1}{1 - w_1'} - \frac{w_2' X_{n-1} + \cdots + w_n' X_1}{1 - w_1'}\|_p$$

$$\leq (1 - w_1')\|\frac{(w_1 - w_1')E[X_n|\mathcal{F}_{n-1}]}{1 - w_1'} + \frac{w_2 X_{n-1} + \cdots + w_n X_1}{1 - w_1'} - \frac{w_2' X_{n-1} + \cdots + w_n' X_1}{1 - w_1'}\|_p$$

$$+(w_1 - w_1')\|\epsilon_n\|_p$$

by Minkowski Inequality.

Hence

$$\|w_1 X_n + w_2 X_{n-1} + \cdots + w_n X_1 - (w_1' X_n + w_2' X_{n-1} + \cdots + w_n' X_1)\|_p \leq (1 - w_1')D_{n-1} + (w_1 - w_1')\|\epsilon_n\|_p$$

So we have

$$D_n \leq \frac{n-1}{n} D_{n-1} + \frac{n-1}{n}\|\epsilon_n\|_p$$

Then by similar argument we can show when $\sum_{n=1}^{\infty} \|\epsilon_n\|_p < \infty$, $D_n \to 0$.

Thus X_n converges in L^p. □

To prove the almost sure convergence of a GSAP, we need the following lemma of random variables.

Lemma 6.0.8. *Suppose $X_n \to X$ in probability and $\sum_1^{\infty} P(|X_n - X| \geq \epsilon) < \infty$ for any $\epsilon > 0$. Then $X_n \to X$ a.s.*

Proof. By Borel-Cantelli Lemma, $P(|X_n - X| \geq \epsilon, i.o.) = 0$ so $X_n \to X$ a.s. □

Theorem 6.0.9. *(Almost Sure Convergence)*

Let $\{X_n\}$ be a GSAP for the array T and suppose $\epsilon_n = X_n - E[X_n|\mathcal{F}_{n-1}]$ satisfies $\sum nE|\epsilon_n|^2 < \infty$. Then $\{X_n\}$ converges almost surely.

Proof. We know $\{X_n\}$ converges in L^2 by Theorem 6.0.6 and let the limit be X. By the Markov inequality, $P(|X_n - X| \geq \epsilon) \leq \frac{E|X_n - X|^2}{\epsilon^2}$, so it suffices to show $E|X_n - X|^2$ is summable by the previous lemma. Let $n \geq m$:

$$\begin{aligned} E|X_n - X_m|^2 &= E|X_m - E[X_n|\mathcal{F}_m]|^2 + E|X_n - E[X_n|\mathcal{F}_m]|^2 \\ &\leq E|\epsilon_n|^2 + \cdots + E|\epsilon_{m+1}|^2 + (\frac{m-1}{m})^2 E|\epsilon_m|^2 + \cdots + (\frac{2}{m})^2 E|\epsilon_3|^2 + (\frac{2}{m})^2 E|\epsilon_2|^2 \end{aligned}$$

Taking $n \to \infty$ we have

$$E|X_m - X|^2 \leq (\sum_{k=m+1}^{\infty} E|\epsilon_k|^2) + (\frac{m-1}{m})^2 E|\epsilon_m|^2 + \cdots + (\frac{2}{m})^2 E|\epsilon_3|^2 + (\frac{2}{m})^2 E|\epsilon_2|^2.$$

Then

$$\sum_{m=1}^{\infty} E|X_m - X|^2 \leq \sum_{m=1}^{\infty} [(\sum_{k=m+1}^{\infty} E|\epsilon_k|^2) + (\frac{m-1}{m})^2 E|\epsilon_m|^2 + \cdots + (\frac{2}{m})^2 E|\epsilon_3|^2 + (\frac{2}{m})^2 E|\epsilon_2|^2]$$

By Tonelli theorem we can change the order of the summations:

$$\sum_{m=1}^{\infty} E|X_m - X|^2 \leq (\sum_{i=1}^{\infty} \frac{2}{i^2}) E|\epsilon_2|^2 + \sum_{m=3}^{\infty} (m - 2 + \sum_{i=m-1}^{\infty} \frac{(m-1)^2}{i^2}) E|\epsilon_m|^2$$

since $\sum_{i=m-1}^{\infty} \frac{1}{i^2} < \int_{m-2}^{\infty} \frac{1}{x^2} dx = \frac{1}{m-2}$,

$$\begin{aligned} \sum_{m=1}^{\infty} E|X_m - X|^2 &< (\sum_{i=1}^{\infty} \frac{2}{i^2}) E|\epsilon_2|^2 + \sum_{m=3}^{\infty} (m - 2 + \frac{(m-1)^2}{m-2}) E|\epsilon_m|^2 \\ &< (\sum_{i=1}^{\infty} \frac{2}{i^2}) E|\epsilon_2|^2 + \sum_{m=3}^{\infty} 2m E|\epsilon_m|^2 < \infty. \end{aligned}$$

So $X_n \to X$ almost surely. □

Bibliography

[1] Doob, J. L. *Stochastic Processes*. John Wiley & Sons, 1953.

[2] Fisk, D. L. Quasi-martingales. *Transactions of the American Mathematical Society 120* (1965), 369–389.

[3] Hammersley, J. M. Harnesses. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 3* (1967), 89–117.

[4] Mansuy, R. The origins of the word "martingale". *Electronic Journal for History of Probability and Statistics 5* (2009), 1–10.

[5] Merca, M. A note on determinat of a toeplitz-hessenberg matrix. *Special Matrices 1* (2013), 10–16.

[6] Tomkins, R. J. Martingale generalizations and preservation of martingale properties. *The Canadian Journal of Statistics 12* (1984), 99–106.

[7] Williams, D. *Probability with Martingales*. Cambridge University Press Beijing World Pubishing, 2008.