EXPECTATION, CONDITIONAL EXPECTATION AND MARTINGALES IN LOCAL FIELDS

STEVEN N. EVANS AND TYE LIDMAN

ABSTRACT. We investigate a possible definition of expectation and conditional expectation for random variables with values in a local field such as the p-adic numbers. We define the expectation by analogy with the observation that for real-valued random variables in L^2 the expected value is the orthogonal projection onto the constants. Previous work has shown that the local field version of L^∞ is the appropriate counterpart of L^2, and so the expected value of a local field-valued random variable is defined to be its "projection" in L^∞ onto the constants. Unlike the real case, the resulting projection is not typically a single constant, but rather a ball in the metric on the local field. However, many properties of this expectation operation and the corresponding conditional expectation mirror those familiar from the real-valued case; for example, conditional expectation is, in a suitable sense, a contraction on L^∞ and the tower property holds. We also define the corresponding notion of martingale, show that several standard examples of martingales (for example, sums or products of suitable independent random variables or "harmonic" functions composed with Markov chains) have local field analogues, and obtain versions of the optional sampling and martingale convergence theorems.

1. INTRODUCTION

Expectation and conditional expectation of real-valued random variables (or, more generally, Banach space-valued random variables) and the corresponding notion of martingale are fundamental objects of probability theory. In this paper we investigate whether there are analogous notions for random variables with values in a local field (that is, a locally compact, non-discrete, totally disconnected, topological field) – a setting that shares the linear structure which underpins many of the properties of the classical entities.

The best known example of a local field is the field of *p-adic numbers* for some positive prime p. This field is defined as follows. We can write any non-zero rational number $r \in \mathbb{Q}\backslash\{0\}$ uniquely as $r = p^s(a/b)$, with

SNE supported in part by NSF grant DMS-0405778.

a, b, and s integers, where a and b are not divisible by p. Set $|r| = p^{-s}$. If we set $|0| = 0$, then the map $|\cdot|$ has the properties:

$$
\begin{aligned}
|x| &= 0 \Leftrightarrow x = 0 \\
|xy| &= |x||y| \\
|x+y| &\le |x| \vee |y|.
\end{aligned} \tag{1}
$$

The map $(x, y) \mapsto |x - y|$ defines a metric on $\mathbb{Q}$ and we denote the completion of $\mathbb{Q}$ in this metric by $\mathbb{Q}_p$. The field operations on $\mathbb{Q}$ extend continuously to make $\mathbb{Q}_p$ a topological field called the *p-adic numbers*. The map $|\cdot|$ also extends continuously and the extension has properties (1).

The closed unit ball around 0, $\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x| \le 1\}$, is the closure in $\mathbb{Q}_p$ of the integers $\mathbb{Z}$, and is thus a ring (this is also apparent from (1)), called the *p-adic integers*. As $\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x| < p\}$, the set $\mathbb{Z}_p$ is also open. Any other ball around 0 is of the form $\{x \in \mathbb{Q}_p : |x| \le p^{-k}\} = p^k\mathbb{Z}_p$ for some integer k.

Every local field is either a finite algebraic extension of the p-adic number field for some prime p or a finite algebraic extension of the *p-series field*; that is, the field of formal Laurent series with coefficients drawn from the finite field with p elements.) A locally compact, non-discrete, topological field that is not totally disconnected is necessarily either the real or the complex numbers.

From now on, we let $\mathbb{K}$ be a fixed local field. Good general reference for the properties of local fields and analysis on them are [Sch84, vR78, Tai75]. The following are the properties we need.

There is a real-valued mapping $x \mapsto |x|$ on $\mathbb{K}$ called the non-archimedean valuation with the properties (1). The third of these properties is the *ultrametric inequality* or the *strong triangle inequality*. The map $(x, y) \mapsto |x - y|$ on $\mathbb{K} \times \mathbb{K}$ is a metric on $\mathbb{K}$ which gives the topology of $\mathbb{K}$. A consequence of of the strong triangle inequality is that if $|x| \ne |y|$, then $|x + y| = |x| \vee |y|$. This latter result implies that for every "triangle" $\{x, y, z\} \subset \mathbb{K}$ we have that at least two of the lengths $|x - y|$, $|x - z|$, $|y - z|$ must be equal and is therefore often called the *isosceles triangle property*.

The valuation takes the values $\{q^k : k \in \mathbb{Z}\} \cup \{0\}$, where $q = p^c$ for some prime p and positive integer c (so that for $\mathbb{K} = \mathbb{Q}_p$ we have $c = 1$). Write $\mathbb{D}$ for $\{x \in \mathbb{K} : |x| \le 1\}$ (so that $\mathbb{D} = \mathbb{Z}_p$ when $\mathbb{K} = \mathbb{Q}_p$). Fix $\rho \in \mathbb{K}$ so that $|\rho| = q^{-1}$. Then

$$\rho^k\mathbb{D} = \{x : |x| \le q^{-k}\} = \{x : |x| < q^{-(k-1)}\}$$

for each $k \in \mathbb{Z}$ (so that for $\mathbb{K} = \mathbb{Q}_p$ we could take $\rho = p$). The set $\mathbb{D}$ is the unique maximal compact subring of $\mathbb{K}$ (the *ring of integers* of $\mathbb{K}$). Every ball in $\mathbb{K}$ is of the form $x + \rho^k\mathbb{D}$ for some $x \in \mathbb{D}$ and $k \in \mathbb{Z}$. If $B = x + \rho^k\mathbb{D}$ and $C = y + \rho^\ell\mathbb{D}$ are two such balls, then

- $B \cap C = \emptyset$, if $|x - y| > q^{-k} \vee q^{-\ell}$,
- $B \subseteq C$, if $|x - y| \vee q^{-k} \le q^{-\ell}$,
- $C \subseteq B$, if $|x - y| \vee q^{-\ell} \le q^{-k}$.

In particular, if $q^{-k} = q^{-\ell}$, then either $B \cap C = \emptyset$ or $B = C$, depending on whether or not $|x - y| > q^{-k} = q^{-\ell}$ or $|x - y| \le q^{-k} = q^{-\ell}$.

We have shown in a sequence papers [Eva89, Eva91, Eva93, Eva95, Eva01b, Eva01a, Eva02, Eva06] that the natural analogues on $\mathbb{K}$ of the centered Gaussian measures on $\mathbb{R}$ are the normalized restrictions of Haar measure on the additive group of $\mathbb{K}$ to the compact the balls $\rho^k\mathbb{D}$ and the point mass at 0. There is a significant literature on probability on the p-adics and other local fields. The above papers contain numerous references to this work, much of which concerns Markov processes taking values in local fields. There are also extensive surveys of the literature in the books [Khr97, Koc01, KN04].

It is not immediately clear how one should approach defining the expectation of a local field valued random variable X. Even if X only takes a finite number of values $\{x_1, x_2, \dots, x_n\}$, then the object $\sum_k x_k \mathbb{P}\{X = x_k\}$ doesn't make any sense because $x_k \in \mathbb{K}$ whereas $\mathbb{P}\{X = x_k\} \in \mathbb{R}$. However, it is an elementary fact that if T is a real-valued random variable with $\mathbb{E}[T^2] < \infty$, then $c \mapsto \mathbb{E}[(T - c)^2]$ is uniquely minimized by $c = \mathbb{E}[T]$. Of course, since this observation already uses the notion of expectation it does not lead to an alternative way of defining the expected value of a real-valued random variable. Fortunately, we can do something similar, but non-circular, in the local field case.

Fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. By a $\mathbb{K}$-valued random variable, we mean a measurable map from Ω equipped with $\mathcal{F}$ into $\mathbb{K}$ equipped with its Borel σ-field. Let L^∞ be the space of $\mathbb{K}$-valued random variables X that satisfy $\|X\|_\infty := \operatorname{ess\,sup} |X| < \infty$. It is clear that L^∞ is a vector space over $\mathbb{K}$. If we identify two random variables as being equal when they are equal almost surely, then

$$
\begin{aligned}
\|X\|_\infty = 0 &\Leftrightarrow X = 0 \\
\|cX\|_\infty &= |c|\|X\|_\infty, \quad c \in \mathbb{K}, \\
\|X + Y\|_\infty &\le \|X\|_\infty \vee \|Y\|_\infty.
\end{aligned}
$$

The map $(X, Y) \mapsto \|X - Y\|_\infty$ defines a metric on L^∞ (or, more correctly, on equivalence classes under the relation of equality almost everywhere), and L^∞ is complete in this metric. Hence L^∞ is an instance of a Banach algebra over $\mathbb{K}$.

It is apparent from the papers on analogues of Gaussian measures cited above that L^∞ is the natural local field counterpart of the real Hilbert space L^2. In particular, there is a natural notion of orthogonality on L^∞ (albeit one which does not come from an inner product structure).

Definition 1.1. Given $X \in L^\infty$, set $\varepsilon(X) = \inf\{\|X - c\|_\infty : c \in \mathbb{K}\}$. The *expectation* of the $\mathbb{K}$-valued random variable X is the subset of $\mathbb{K}$ given by

$$\mathbb{E}[X] := \{c \in \mathbb{K} : \|X - c\|_\infty = \varepsilon(X)\} = \{c \in \mathbb{K} : \|X - c\|_\infty \leq \varepsilon(X)\}.$$

We show in Section 2 that $\mathbb{E}[X]$ is non-empty. Note that if $c' \in \mathbb{E}[X]$ and $c'' \in \mathbb{K}$ is such that $|c'' - c'| \leq \varepsilon(X)$, then, by the strong triangle inequality, $c'' \in \mathbb{E}[X]$. Thus $\mathbb{E}[X]$ is a (closed) ball in $\mathbb{K}$ (where we take a single point as being a ball).

Observe that we use the same notation for expectation of $\mathbb{K}$-valued and $\mathbb{R}$-valued random variables. This should cause no confusion: we either indicate explicitly whether a random variable has values in $\mathbb{K}$ or $\mathbb{R}$, or this will be clear from context.

The outline of the rest of the paper is the following. We show in Section 2 that the expected value of a random variable in L^∞ is non-empty, remark on some of the properties of the expectation operator, and motivate the definition of conditional expectation by considering the situation where the conditioning σ-field is finitely generated or, more generally, has an associated regular conditional probability. The appropriate definition of the conditional expectation of $X \in L^\infty$ given a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ is not, as one might first imagine, the L^∞ projection of X onto $L^\infty(\mathcal{G})$ (:= the subspace of L^∞ consisting of $\mathcal{G}$-measurable random variables). For this reason, we need to do some preparatory work in Sections 3 and 4 before finally presenting the construction of conditional expectation in Section 5 and describing its elementary properties in Section 6. We establish an analogue of the "tower property" in Section 7 and obtain a counterpart of the fact for classical conditional expectation that conditioning is a contraction on L^2 (both of these results need to be suitably interpreted due to the conditional expectation being typically a set of random variables rather than a single one). We introduce the associated notion of martingale in Section 9 and observe that several of the classical examples of martingales have local field analogues. We develop counterparts of the optional sampling theorem and martingale convergence theorem in Sections 10 and 11, respectively.

Note: We adopt the convention that all equalities and inequalities between random variables should be interpreted as holding $\mathbb{P}$-almost surely.

2. Expectation

Theorem 2.1. *The expectation of a random variable $X \in L^\infty$ is non-empty. It is the smallest closed ball in $\mathbb{K}$ that contains* $\mathrm{supp} X$ *(the closed support of X).*

Proof. By the strong triangle inequality $\|X - c\|_\infty \le \|X\|_\infty \vee |c|$, and $\|X - c\|_\infty = |c|$ for $|c| > \|X\|_\infty$. Therefore, the infimum of $c \mapsto \|X - c\|_\infty$ over all $c \in \mathbb{K}$ is the same as the infimum over $\{c \in \mathbb{K} : |c| \le \|X\|_\infty\}$ and any point $c \in \mathbb{K}$ at which the infimum of is achieved must necessarily satisfy $|c| \le \|X\|_\infty$. That is, $\varepsilon(X) = \inf\{\|X - c\|_\infty : |c| \le \|X\|_\infty\}$ and $\mathbb{E}[X] = \{c : |c| \le \|X\|_\infty,\ \|X - c\|_\infty = \varepsilon(X)\}$.

Again by the strong triangle inequality, the function $c \mapsto \|X - c\|_\infty$ is continuous. Consequently, $\mathbb{E}[X]$ is non-empty as the set of points at which a continuous function on a compact set attains its infimum.

As we observed in the Introduction, $\mathbb{E}[X]$ is a ball of radius (= diameter) $\varepsilon(X)$. If $x \in \mathrm{supp} X$ is not in $\mathbb{E}[X]$ and c is any point in $\mathbb{E}[X]$, then, by the strong triangle inequality, $|x - c| > \varepsilon(X)$ and $\|X - c\|_\infty > \varepsilon(X)$, contradicting the definition of $\mathbb{E}[X]$. Thus $\mathrm{supp} X \subseteq \mathbb{E}[X]$. Hence, if the smallest ball containing $\mathrm{supp} X$ is not $\mathbb{E}[X]$, it must be a ball contained in $\mathbb{E}[X]$ with diameter $r < \varepsilon(X)$. However, if c is any point contained in the smaller ball, then $|x - c| \le r$ for all $x \in \mathrm{supp} X$, contradicting the definition of $\varepsilon(X)$. □

Our notion of expectation shares some of the features of both the mean and the variance of a real-valued variable. Any point in the ball $\mathbb{E}[X]$ is as good a single summary of the "location" of X as any other, whereas the diameter of $\mathbb{E}[X]$ (that is, $\varepsilon(X)$) is a measure of the "spread" of X.

Some properties of $\mathbb{E}[X]$ are immediate. It is easily seen that for constants $k, b \in \mathbb{K}$, $\mathbb{E}[kX + b] = k\mathbb{E}[X] + b$. We do not have complete linearity, however, since $\mathbb{E}[X + Y]$ is only a subset of $\mathbb{E}[X] + \mathbb{E}[Y]$, with equality when X and Y are independent. This follows from the fact that $\mathrm{supp}(X + Y) \subseteq \mathrm{supp} X + \mathrm{supp} Y$, with equality when X and Y are independent. Also, if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. These remarks further support our assertion that

$\mathbb{E}[X]$ combines the properties of the mean and the variance for real-valued random variables.

Define the Hausdorff distance between two subsets A and B of $\mathbb{K}$ to be

$$d_H(A,B) := \sup_{a\in A}\inf_{b\in B}|a-b| \vee \sup_{b\in B}\inf_{a\in A}|b-a|.$$

We know from Theorem 2.1 that $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are balls with diameters $\varepsilon(X)$ and $\varepsilon(Y)$, respectively. We have one of the alternatives $\mathbb{E}[X]=\mathbb{E}[Y]$, $\mathbb{E}[X]\subsetneq\mathbb{E}[Y]$, $\mathbb{E}[Y]\subsetneq\mathbb{E}[X]$, or $\mathbb{E}[X]\cap\mathbb{E}[Y]=\emptyset$. Suppose that $\mathbb{E}[X]\subsetneq\mathbb{E}[Y]$, so that $\mathrm{supp}X\subseteq\mathbb{E}[X]$ and there exists $y\in\mathrm{supp}Y$ such that y is not in the unique ball of diameter $q^{-1}\varepsilon(Y)$ containing $\mathbb{E}[X]$. Then, by the strong triangle inequality, $|x-y|=\varepsilon(Y)$ for all $x\in\mathrm{supp}X$, and so $d_H(\mathrm{supp}X,\mathrm{supp}Y)\geq\varepsilon(Y)=d_H(\mathbb{E}[X],\mathbb{E}[Y])$ in this case. Similar arguments in the other cases show that

$$d_H(\mathbb{E}[X],\mathbb{E}[Y])\leq d_H(\mathrm{supp}X,\mathrm{supp}Y)\leq\|X-Y\|_\infty.$$

This is analogous to the continuity of real-valued expectation with respect to the real L^p norms.

Rather than develop more properties of expectation, we move on to the corresponding definition of conditional expectation because, just as in the real case, expectation is the special case of conditional expectation that occurs when the conditioning σ-field is the trivial σ-field $\{\emptyset,\Omega\}$, and so results for expectation are just special cases of ones for conditional expectation.

In order to motivate the definition of conditional expectation, first consider the special case when the conditioning σ-field $\mathcal{G}\subseteq\mathcal{F}$ is generated by a finite partition $\{A_1,A_2,\ldots,A_n\}$ of Ω. In line with our definition of $\mathbb{E}[X]$, a reasonable definition of $\mathbb{E}[X\,|\,\mathcal{G}]$ would be the set of $\mathcal{G}$-measurable random variables Y such that for each k the common value of $c_k := Y(\omega)$ for $\omega\in A_k$ satisfies

$$\operatorname{ess\,sup}\{|X(\omega)-c_k| : \omega\in A_k\} = \inf_{c\in\mathbb{K}}\operatorname{ess\,sup}\{|X(\omega)-c| : \omega\in A_k\}.$$

Equivalently, suppose we define $\varepsilon(X,\mathcal{G})$ to be the $\mathcal{G}$-measurable, $\mathbb{R}$-valued random variable that takes the value $\inf_{c\in\mathbb{K}}\operatorname{ess\,sup}\{|X(\omega)-c| : \omega\in A_k\}$ on A_k, then $\mathbb{E}[X\,|\,\mathcal{G}]$ is the set of $\mathcal{G}$-measurable random variables Y such that $|X-Y|\leq\varepsilon(X,\mathcal{G})$. Note that $\varepsilon(X,\{\emptyset,\Omega\})=\varepsilon(X)$ and $\mathbb{E}[X\,|\,\{\emptyset,\Omega\}]=\mathbb{E}[X]$.

More generally, suppose that $\mathcal{G}\subseteq\mathcal{F}$ is an arbitrary sub-σ-field and there is an associated regular conditional probability $\mathbb{P}_\mathcal{G}(\omega',d\omega'')$ (such a regular conditional probability certainly exists if $\mathcal{G}$ is finitely generated). In this case, we expect that $\mathbb{E}[X\,|\,\mathcal{G}](\omega')$ should be the expectation of X with respect to the probability measure $\mathbb{P}_\mathcal{G}(\omega',\cdot)$. It is easy to see that if we let $\varepsilon(X, \mathcal{G})$ be the $\mathcal{G}$-measurable random variable such that $\varepsilon(X, \mathcal{G})(\omega')$ is the infimum over $c \in \mathbb{K}$ of the essential supremum of $|X - c|$ with respect to $\mathbb{P}_{\mathcal{G}}(\omega', \cdot)$, then this definition of $\varepsilon(X, \mathcal{G})$ subsumes our previous one for the finitely generated case and our putative definition of $\mathbb{E}[X \mid \mathcal{G}]$ coincides with the set of $\mathcal{G}$-measurable random variables Y such that $|X - Y| \le \varepsilon(X, \mathcal{G})$, thereby also extending the definition for the finitely generated case.

We therefore see that the key to giving a satisfactory general definition of $\mathbb{E}[X \mid \mathcal{G}]$ for an arbitrary sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$ is to find a suitable general definition of $\varepsilon(X, \mathcal{G})$. We tackle this problem in the next three sections.

3. Conditional essential supremum

Definition 3.1. Given a non-negative real-valued random variable S and a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, put

$$\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = \sup_{p \ge 1} \mathbb{E}[S^p \mid \mathcal{G}]^{\frac{1}{p}} = \lim_{p \to \infty} \mathbb{E}[S^p \mid \mathcal{G}]^{\frac{1}{p}}.$$

Lemma 3.2. (i) *Suppose that S is a non-negative real-valued random variable and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Then* $S \le \operatorname{ess\,sup}\{S \mid \mathcal{G}\}$.

(ii) *Suppose that S and $\mathcal{G}$ are as in (i) and T is $\mathcal{G}$-measurable real-valued random variable with $S \le T$. Then* $\operatorname{ess\,sup}\{S \mid \mathcal{G}\} \le T$.

(iii) *Suppose that S' and S'' are non-negative real-valued random variables and $\mathcal{G}$ is a sub-σ-fields of $\mathcal{F}$. Then*

$$\operatorname{ess\,sup}\{S' \vee S'' \mid \mathcal{G}\} = \operatorname{ess\,sup}\{S' \mid \mathcal{G}\} \vee \operatorname{ess\,sup}\{S'' \mid \mathcal{G}\}.$$

Proof. For part (i), we show by separate arguments that the result holds on the events $\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = 0\}$ and $\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} > 0\}$.

First consider what happens on the event $\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = 0\}$. By definition $\mathbb{E}[S \mid \mathcal{G}] \le \operatorname{ess\,sup}\{S \mid \mathcal{G}\}$. Hence

$$\begin{aligned}\mathbb{E}[S\,\mathbf{1}\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = 0\}] &\le \mathbb{E}[S\,\mathbf{1}\{\mathbb{E}[S \mid \mathcal{G}] = 0\}] \\ &= \mathbb{E}[\mathbb{E}[S\,\mathbf{1}\{\mathbb{E}[S \mid \mathcal{G}] = 0\} \mid \mathcal{G}]] \\ &= \mathbb{E}[\,\mathbf{1}\{\mathbb{E}[S \mid \mathcal{G}] = 0\}\mathbb{E}[S \mid \mathcal{G}]] = 0.\end{aligned}$$

Thus $\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = 0\} \subseteq \{S = 0\}$, and $S \le \operatorname{ess\,sup}\{S \mid \mathcal{G}\}$ on the event $\{\operatorname{ess\,sup}\{S \mid \mathcal{G}\} = 0\}$.

Now consider what happens on the event $\{\operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} > 0\}$. Take $\alpha > 1$. Observe for $p \geq 1$ that

$$\begin{aligned}\mathbb{E}[S^p \,|\, \mathcal{G}] &\geq \mathbb{E}[S^p \, \mathbf{1}\{S^p \geq \alpha^p \mathbb{E}[S^p \,|\, \mathcal{G}]\} \,|\, \mathcal{G}] \\ &\geq \mathbb{E}[\alpha^p \mathbb{E}[S^p \,|\, \mathcal{G}] \, \mathbf{1}\{S^p \geq \alpha^p \mathbb{E}[S^p \,|\, \mathcal{G}]\} \,|\, \mathcal{G}] \\ &= \alpha^p \mathbb{E}[S^p \,|\, \mathcal{G}] \, \mathbb{P}\{S^p \geq \alpha^p \mathbb{E}[S^p \,|\, \mathcal{G}] \,|\, \mathcal{G}\}.\end{aligned}$$

Hence, for each $p \geq 1$,

$$\mathbb{P}\{S \geq \alpha \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} \,|\, \mathcal{G}\} \leq \mathbb{P}\{S \geq \alpha \, \mathbb{E}[S^p \,|\, \mathcal{G}]^{\frac{1}{p}} \,|\, \mathcal{G}\} \leq \frac{1}{\alpha^p}$$

on the event $\{\mathbb{E}[S^p \,|\, \mathcal{G}] > 0\}$.

Since $\{\operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} > 0\} \subseteq \bigcup_p \bigcap_{q \geq p} \{\mathbb{E}[S^q \,|\, \mathcal{G}] > 0\}$, we see that $\mathbb{P}\{S \geq \alpha \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} \,|\, \mathcal{G}\} = 0$ on the event on $\{\operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} > 0\}$. As this holds for all $\alpha > 1$, we conclude that $S \leq \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\}$ on the event $\{\operatorname{ess\,sup}\{S \,|\, \mathcal{G}\} > 0\}$, and this completes the proof of part (i).

Part (ii) is immediate from the definition.

Now consider part (iii). We have from part (i) that $S' \leq \operatorname{ess\,sup}\{S' \,|\, \mathcal{G}\}$ and $S'' \leq \operatorname{ess\,sup}\{S'' \,|\, \mathcal{G}\}$. Thus $S' \vee S'' \leq \operatorname{ess\,sup}\{S' \,|\, \mathcal{G}\} \vee \operatorname{ess\,sup}\{S'' \,|\, \mathcal{G}\}$ and hence

$$\operatorname{ess\,sup}\{S' \vee S'' \,|\, \mathcal{G}\} \leq \operatorname{ess\,sup}\{S' \,|\, \mathcal{G}\} \vee \operatorname{ess\,sup}\{S'' \,|\, \mathcal{G}\}$$

by part (ii). On the other hand, because $S' \leq S' \vee S''$ and $S'' \leq S' \vee S''$, it follows that $\operatorname{ess\,sup}\{S' \,|\, \mathcal{G}\} \leq \operatorname{ess\,sup}\{S' \vee S'' \,|\, \mathcal{G}\}$ and $\operatorname{ess\,sup}\{S'' \,|\, \mathcal{G}\} \leq \operatorname{ess\,sup}\{S' \vee S'' \,|\, \mathcal{G}\}$. Therefore

$$\operatorname{ess\,sup}\{S' \,|\, \mathcal{G}\} \vee \operatorname{ess\,sup}\{S'' \,|\, \mathcal{G}\} \leq \operatorname{ess\,sup}\{S' \vee S'' \,|\, \mathcal{G}\}.$$

□

Corollary 3.3. *Suppose that S is a non-negative real-valued random variable and $\mathcal{G} \subseteq \mathcal{H}$ are sub-σ-fields of $\mathcal{F}$. Then* $\operatorname{ess\,sup}\{S \,|\, \mathcal{H}\} \leq \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\}$.

Proof. From Lemma 3.2(i), $S \leq \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\}$. Applying Lemma 3.2(ii) with $\mathcal{G}$ replaced by $\mathcal{H}$ and $T = \operatorname{ess\,sup}\{S \,|\, \mathcal{G}\}$ gives the result. □

Let $\{\mathcal{F}_n\}_{n=0}^\infty$ be a filtration (that is, a non-decreasing sequence of sub-σ-fields of $\mathcal{F}$). Recall that a random variable T with values in $\{0, 1, 2, \ldots\}$ is a stopping time for the filtration if $\{T = n\} \in \mathcal{F}_n$ for all n. Recall also that if T is a stopping time, then the associated σ-field $\mathcal{F}_T$ is the collection of events A such that $A \cap \{T = n\} \in \mathcal{F}_n$ for all n.

Lemma 3.4. *Suppose that S is a non-negative real-valued random variable, $\{\mathcal{F}_n\}_{n=0}^\infty$ is a filtration of sub-σ-fields of $\mathcal{F}$, and T is a stopping* *time. Then*

$$\begin{aligned}\operatorname{ess\,sup}\{S\,\mathbf{1}\{T=n\}\mid\mathcal{F}_T\} &= \mathbf{1}\{T=n\}\operatorname{ess\,sup}\{S\mid\mathcal{F}_T\}\\ &= \mathbf{1}\{T=n\}\operatorname{ess\,sup}\{S\mid\mathcal{F}_n\} = \operatorname{ess\,sup}\{S\,\mathbf{1}\{T=n\}\mid\mathcal{F}_n\}\end{aligned}$$

for all n.

Proof. This follows immediately from the definition of the conditional essential supremum and the fact that if U is a non-negative real-valued random variable, then $\operatorname{ess\,sup}\{U|\mathcal{F}_T\} = \operatorname{ess\,sup}\{U|\mathcal{F}_n\}$ on the event $\{T=n\}$ (see, for example, Proposition II-1-3 of [Nev75]). □

4. Conditional L^∞ norm

Definition 4.1. Given $X \in L^\infty$ and a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, put

$$\|X\|_{\mathcal{G}} := \operatorname{ess\,sup}\{|X| \mid \mathcal{G}\}.$$

Notation 4.2. Given $A \in \mathcal{F}$, the $\mathbb{K}$-valued random variable $\mathbf{1}_A$ is given by

$$\mathbf{1}_A(\omega) = \begin{cases} 1_{\mathbb{K}}, & \text{if } \omega \in A,\\ 0_{\mathbb{K}}, & \text{otherwise,}\end{cases}$$

where $1_{\mathbb{K}}$ and $0_{\mathbb{K}}$ are, respectively, the multiplicative and additive identity elements of $\mathbb{K}$. We continue to use this same notation to also denote the analogously defined real-valued indicator random variable, but this should cause no confusion as the meaning will be clear from the context.

Lemma 4.3. *Fix a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.*

(i) *If $W \in L^\infty(\mathcal{G})$ and $X \in L^\infty$, then $\|WX\|_{\mathcal{G}} = |W|\,\|X\|_{\mathcal{G}}$.*
(ii) *If $X, Y \in L^\infty$ are such that $\mathbb{P}(\{X \neq Y\} \cap A) = 0$ for some $A \in \mathcal{G}$, then $\mathbb{P}(\{\|X\|_{\mathcal{G}} \neq \|Y\|_{\mathcal{G}}\} \cap A) = 0$.*
(iii) *If $X_1, X_2, \ldots \in L^\infty$ and $A_1, A_2, \ldots \in \mathcal{G}$ are pairwise disjoint, then*

$$\left\|\sum_i X_i\,\mathbf{1}_{A_i}\right\|_{\mathcal{G}} = \sum_i \mathbf{1}_{A_i}\|X_i\|_{\mathcal{G}}.$$

(iv) *If $X, Y \in L^\infty$, then*

$$\|X+Y\|_{\mathcal{G}} \le \|X\|_{\mathcal{G}} \vee \|Y\|_{\mathcal{G}}.$$

Proof. Part (i) follows immediately from the definition. Part (ii) follows from part (i): since $X\,\mathbf{1}_A = Y\,\mathbf{1}_A$ by assumption,

$$\mathbf{1}_A\|X\|_{\mathcal{G}} = \|X\,\mathbf{1}_A\|_{\mathcal{G}} = \|Y\,\mathbf{1}_A\|_{\mathcal{G}} = \mathbf{1}_A\|Y\|_{\mathcal{G}}.$$

Part (iii) follows from parts (i) and (ii): for any of the events A_j,

$$\begin{aligned}\mathbf{1}_{A_j}\sum_i \mathbf{1}_{A_i}\|X_i\|_{\mathcal{G}} &= \mathbf{1}_{A_j}\|X_j\|_{\mathcal{G}} = \|\,\mathbf{1}_{A_j}X_j\|_{\mathcal{G}} \\ &= \left\|\mathbf{1}_{A_j}\sum_i \mathbf{1}_{A_i}X_i\right\|_{\mathcal{G}} = \mathbf{1}_{A_j}\left\|\sum_i \mathbf{1}_{A_i}X_i\right\|_{\mathcal{G}},\end{aligned}$$

and, similarly, $\sum_i \mathbf{1}_{A_i}\|X_i\|_{\mathcal{G}} = \|\sum_i \mathbf{1}_{A_i}X_i\|_{\mathcal{G}}$ on $\Omega\setminus(\bigcup_i A_i)$.

Part (iv) is an immediate consequence of Lemma 3.2(iii). However, there is also the following alternative, more elementary proof. Note first that $\|X^r\|_{\mathcal{G}} = \|X\|_{\mathcal{G}}^r$ for any $r>0$ because

$$\lim_{p\to\infty}\mathbb{E}[|X|^{rp}\,|\,\mathcal{G}]^{\frac{1}{p}} = \lim_{q\to\infty}\mathbb{E}[|X|^q\,|\,\mathcal{G}]^{\frac{r}{q}} = (\lim_{q\to\infty}\mathbb{E}[|X|^q\,|\,\mathcal{G}]^{\frac{1}{q}})^r.$$

Thus, from Jensen's inequality and the observation that $(x+y)^s \le (x^s+y^s)$ for $0\le s\le 1$,

$$\begin{aligned}\|X+Y\|_{\mathcal{G}} &= \lim_{p\to\infty}\mathbb{E}[|X+Y|^p\,|\,\mathcal{G}]^{\frac{1}{p}} \\ &\le \lim_{p\to\infty}\mathbb{E}[|X|^p\vee|Y|^p\,|\,\mathcal{G}]^{\frac{1}{p}} \\ &= \lim_{p\to\infty}\mathbb{E}[\lim_{r\to\infty}(|X|^{pr}+|Y|^{pr})^{\frac{1}{r}}\,|\,\mathcal{G})^{\frac{1}{p}} \\ &\le \lim_{p,r\to\infty}(\mathbb{E}[|X|^{pr}\,|\,\mathcal{G}]+\mathbb{E}[|Y|^{pr}\,|\,\mathcal{G}])^{\frac{1}{pr}} \\ &\le \lim_{p,r\to\infty}(\mathbb{E}[|X|^{rp}\,|\,\mathcal{G}]^{\frac{1}{p}}+\mathbb{E}[|Y|^{rp}\,|\,\mathcal{G}]^{\frac{1}{p}})^{\frac{1}{r}}. \\ &= \lim_{r\to\infty}(\|X\|_{\mathcal{G}}^r+\|Y\|_{\mathcal{G}}^r)^{\frac{1}{r}} \\ &= \|X\|_{\mathcal{G}}\vee\|Y\|_{\mathcal{G}}.\end{aligned}$$

□

The following result is immediate from Corollary 3.3.

Lemma 4.4. *Suppose that $X\in L^\infty$ and $\mathcal{G}\subseteq\mathcal{H}$ are sub-σ-fields of $\mathcal{F}$. Then $\|X\|_{\mathcal{H}}\le\|X\|_{\mathcal{G}}$.*

The following result is immediate from Lemma 3.4.

Lemma 4.5. *Suppose that $X\in L^\infty$, $\{\mathcal{F}_n\}_{n=0}^\infty$ is a filtration of sub-σ-fields of $\mathcal{F}$, and T is a stopping time. Then*

$$\begin{aligned}\|X\,\mathbf{1}\{T=n\}\|_{\mathcal{F}_T} &= \mathbf{1}\{T=n\}\,\|X\|_{\mathcal{F}_T} \\ &= \mathbf{1}\{T=n\}\,\|X\|_{\mathcal{F}_n} = \|X\,\mathbf{1}\{T=n\}\|_{\mathcal{F}_n}\end{aligned}$$

for all n.

5. Construction of Conditional Expectation

Definition 5.1. Given $X \in L^\infty$ and a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$, set

$$\mathbb{E}[X \,|\, \mathcal{G}] := \{Y \in L^\infty(\mathcal{G}) : \|X - Y\|_\mathcal{G} \le \|X - Z\|_\mathcal{G} \text{ for all } Z \in L^\infty(\mathcal{G})\}.$$

Remark 5.2. Before showing that $\mathbb{E}[X \,|\, \mathcal{G}]$ is non-empty, we comment on a slight subtlety in the definition. One way of thinking of our definition of $\mathbb{E}[X]$ as the set of $c \in \mathbb{K}$ for which $\|X - c\|_\infty$ is minimal, is that $\mathbb{E}[X]$ is the set of projections of X onto $\mathbb{K} \equiv L^\infty(\{\emptyset, \Omega\})$. A possible definition of $\mathbb{E}[X \,|\, \mathcal{G}]$ might therefore be the analogous set of projections of X onto $L^\infty(\mathcal{G})$, that is, the set of $Y \in L^\infty(\mathcal{G})$ that minimize $\|X - Y\|_\infty$. This definition is **not** equivalent to ours. For example, suppose that Ω consists of the three points $\{\alpha, \beta, \gamma\}$, $\mathcal{F}$ consists of all subsets of Ω, $\mathbb{P}$ assigns positive mass to each point of Ω, $\mathcal{G} = \sigma\{\{\alpha, \beta\}, \{\gamma\}\}$, and X is given by $X(\alpha) = 1_\mathbb{K}$, $X(\beta) = 0_\mathbb{K}$, and $X(\gamma) = 0_\mathbb{K}$. Consider $Y \in L^\infty(\mathcal{G})$, so that $Y(\alpha) = Y(\beta) = c$ and $Y(\gamma) = d$ for some $c, d \in \mathbb{K}$. In order that $Y \in \mathbb{E}[X \,|\, \mathcal{G}]$ according to our definition, c and d must be chosen to minimize both $|1_\mathbb{K} - c| \vee |0_\mathbb{K} - c|$ and $|0_\mathbb{K} - d|$. By the strong triangle inequality, $|1_\mathbb{K} - c| \vee |0_\mathbb{K} - c|$ is minimized by any c with $|c| \le 1$, with the corresponding minimal value being 1. Of course, $|0_\mathbb{K} - d|$ is minimized by the unique value $d = 0_K$. On the other hand, in order that Y is a projection of X onto $L^\infty(\mathcal{G})$, the points c and d must be chosen to minimize $|1_\mathbb{K} - c| \vee |0_\mathbb{K} - c| \vee |0_\mathbb{K} - d|$, and this is accomplished as long as $|c| \le 1$ and $|d| \le 1$. We don't belabor the point in what follows, but several of the natural counterparts of standard results for classical conditional expectation that we show hold for our definition fail to hold for the "projection" definition.

The following lemma is used below to show that $\mathbb{E}[X \,|\, \mathcal{G}]$ is non-empty.

Lemma 5.3. *Suppose that $X \in L^\infty$ is not $0_\mathbb{K}$ almost surely, and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Set $q^{-N} = \|X\|_\infty$. Then there exist disjoint events $A_0, A_1, \ldots \in \mathcal{G}$ and random variables $Y_0, Y_1, \ldots \in L^\infty(\mathcal{G})$ with the following properties:*

1. *On the event A_n, $\|X - Z\|_\mathcal{G} \ge q^{-(N+n)}$ for every $Z \in L^\infty(\mathcal{G})$.*
2. *On the event A_n, $\|X - Y_n\|_\mathcal{G} = q^{-(N+n)}$. and*
3. *On the event $\Omega \setminus \bigcup_{k=1}^n A_k$, $\|X - Y_n\|_\mathcal{G} \le q^{-(N+n+1)}$*
4. *On the event $\bigcup_{k=1}^n A_k$, $Y_p = Y_n$ for any $p > n$.*
5. *The event $\bigcup_{k=1}^\infty A_k$ has probability one.*

Proof. Suppose without loss of generality that $\|X\|_\infty = 1$, so that $N = 0$. Set $\mathbf{Z_0} := \{Z \in L^\infty(\mathcal{G}) : \|X - Z\|_\infty \le 1\}$. Note that the constant 0 belongs to $\mathbf{Z_0}$ and so this set is non-empty. Put $\delta_0 := \inf_{Z\in\mathbf{Z_0}} \mathbb{P}\{\|X-Z\|_{\mathcal{G}}=1\}$.

Choose $Z_{0,1}, Z_{0,2}, \ldots \in \mathbf{Z_0}$ with

$$\lim_{m\to\infty} \mathbb{P}\{\|X-Z_{0,m}\|_{\mathcal{G}}=1\} = \delta_0.$$

Define $Z'_{0,1}, Z'_{0,2}, \ldots$ inductively by setting $Z'_{0,1} := Z_{0,1}$ and

$$Z'_{0,m+1}(\omega) := \begin{cases} Z'_{0,m}(\omega), & \text{if } \|X-Z'_{0,m}\|_{\mathcal{G}}(\omega) \le \|X-Z_{0,m+1}\|_{\mathcal{G}}(\omega), \\ Z_{0,m+1}(\omega), & \text{if } \|X-Z'_{0,m}\|_{\mathcal{G}}(\omega) > \|X-Z_{0,m+1}\|_{\mathcal{G}}(\omega). \end{cases}$$

Note that the events $B_{0,m} := \{\|X-Z'_{0,m}\|_{\mathcal{G}}=1\}$ are decreasing and the $B_{0,m}$ are contained in the event $\{\|X-Z_{0,m}\|_{\mathcal{G}}=1\}$. Hence the event $A_0 := \lim_{m\to\infty} B_{0,m} = \bigcap_{m=1}^{\infty} B_{0,m}$ has probability δ_0.

Define Y_0 by

$$Y_0(\omega) := \begin{cases} Z'_{m,1}(\omega), & \text{if } \omega \in (\Omega\setminus B_{0,1}) \cup A_0, \\ Z'_{0,m}(\omega), & \text{if } \omega \in (\Omega\setminus B_{0,m}) \setminus (\Omega\setminus B_{0,m-1}),\ m\ge 2. \end{cases}$$

It is clear that $\|X-Y_0\|_{\mathcal{G}}=1$ on the event A_0 and $\|X-Y_0\|_{\mathcal{G}} \le q^{-1}$ on the event $\Omega\setminus A_0$. Moreover, if there existed $V \in L^\infty(\mathcal{G})$ with

$$\mathbb{P}(\{\|X-V\|_{\mathcal{G}} \le q^{-1}\} \cap A_0) > 0,$$

then we would have the contradiction that $W \in \mathbf{Z}_0$ defined by

$$W(\omega) = \begin{cases} Y_0(\omega), & \text{if } \|X-Y_0\|_{\mathcal{G}}(\omega) \le \|X-V\|_{\mathcal{G}}(\omega), \\ V(\omega), & \text{if } \|X-Y_0\|_{\mathcal{G}}(\omega) > \|X-V\|_{\mathcal{G}}(\omega) \end{cases}$$

would satisfy $\mathbb{P}\{\|X-W\|_{\mathcal{G}}=1\} < \delta_0$.

Now suppose that $A_0, \ldots A_{n-1}$ and $Y_0, \ldots, Y_{n-1}$ have been constructed with the requisite properties. If $\mathbb{P}(\Omega\setminus\bigcup_{k=1}^{n-1}) = 0$, then take $A_n = \emptyset$ and $Y_n = Y_{n-1}$ (recall that we are interpreting all equalities and inequalities as holding $\mathbb{P}$-a.s.) Otherwise, set

$$\mathbf{Z_n} := \Big\{ Z \in L^\infty(\mathcal{G}) : Z = Y_{n-1} \text{ on } \bigcup_{k=1}^{n-1} A_k \\ \text{and } |X-Z| \le q^{-n} \text{ on } \Omega \setminus \bigcup_{k=1}^{n-1} A_k \Big\}.$$

Note that Y_{n-1} belongs to $\mathbf{Z_n}$. Put $\delta_n := \inf_{Z\in\mathbf{Z_n}} \mathbb{P}\{\|X-Z\|_{\mathcal{G}} = q^{-n}\}$. An argument very similar to the above with $\mathbf{Z_n}$ and δ_n replacing $\mathbf{Z_0}$ and δ_0 establishes the existence of A_n and Y_n with the desired properties. □

Theorem 5.4. *Given $X \in L^\infty$ and a sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, the conditional expectation $\mathbb{E}[X \,|\, \mathcal{G}]$ is nonempty.*

Proof. If X is $0_\mathbb{K}$ almost surely, then $\mathbb{E}[X \,|\, \mathcal{G}] = \{0_\mathbb{K}\}$. Otherwise, let $A_0, A_1, \ldots \in \mathcal{G}$ and $Y_0, Y_1, \ldots \in L^\infty(\mathcal{G})$ be as in Lemma 5.3. Then Y defined by $Y(\omega) = Y_n(\omega)$ for $\omega \in A_n$ belongs to $\mathbb{E}[X \,|\, \mathcal{G}]$. □

6. Elementary Properties of Conditional Expectation

Proposition 6.1. *Fix a sub-σ-field $\mathcal{G} \subseteq \mathcal{F}$.*

(i) *Suppose that $X \in L^\infty(\mathcal{G})$ and $Y \in L^\infty$. Then*
$$\mathbb{E}[XY \,|\, \mathcal{G}] = X\,\mathbb{E}[Y \,|\, \mathcal{G}].$$
and
$$\mathbb{E}[X + Y \,|\, \mathcal{G}] = X + \mathbb{E}[Y \,|\, \mathcal{G}].$$

(ii) *If $X, Y \in L^\infty$ are such that $\mathbb{P}(\{X \neq Y\} \cap A) = 0$ for some $A \in \mathcal{G}$, then $\mathbf{1}_A \mathbb{E}[X \,|\, \mathcal{G}] = \mathbf{1}_A \mathbb{E}[Y \,|\, \mathcal{G}]$.*

(iii) *If $X_1, X_2, \ldots \in L^\infty$ and $A_1, A_2, \ldots \in \mathcal{G}$ are pairwise disjoint, then*
$$\mathbb{E}\left[\sum_i X_i\,\mathbf{1}_{A_i} \,\Big|\, \mathcal{G}\right] = \sum_i \mathbf{1}_{A_i}\mathbb{E}[X_i \,|\, \mathcal{G}].$$

Proof. Consider part (i). We first show the inclusion $\mathbb{E}[XY \,|\, \mathcal{G}] \subseteq X\mathbb{E}[Y \,|\, \mathcal{G}]$.

Consider $Z \in \mathbb{E}[XY \,|\, \mathcal{G}]$. Choose some $V \in \mathbb{E}[Y \,|\, \mathcal{G}]$ and set $W = (Z/X)\,\mathbf{1}\{X \neq 0\} + V\,\mathbf{1}\{X = 0\} \in L^\infty(\mathcal{G})$. Note that $\mathbb{P}\{Z \neq 0,\, X = 0\} = 0$ and hence $XW = Z$, because otherwise we would have the contradiction $\|XY - Z\,\mathbf{1}\{X \neq 0\}\|_\mathcal{G} \le \|XY - Z\|_\mathcal{G}$ and $\mathbb{P}\{\|XY - Z\,\mathbf{1}\{X \neq 0\}\|_\mathcal{G} < \|XY - Z\|_\mathcal{G}\} > 0$ by Lemma 4.3(ii).

We need to show that $W \in \mathbb{E}[Y \,|\, \mathcal{G}]$. Consider $U \in L^\infty(\mathcal{G})$. By Lemma 4.3(ii) and the assumption that $V \in \mathbb{E}[Y \,|\, \mathcal{G}]$,
$$\|Y - W\|_\mathcal{G} = \|Y - V\|_\mathcal{G} \le \|Y - U\|_\mathcal{G}$$
on the event $\{X = 0\}$. Also, $\|XY - Z\|_\mathcal{G} \le \|XY - XU\|_\mathcal{G}$ by the assumption that $Z \in \mathbb{E}[XY \,|\, \mathcal{G}]$, and so, by Lemma 4.3(i)+(ii)
$$\begin{aligned}\|Y - W\|_\mathcal{G} &= \|Y - Z/X\|_\mathcal{G} = |X|^{-1}\|XY - Z\|_\mathcal{G} \\ &\le |X|^{-1}\|XY - XU\|_\mathcal{G} = \|Y - U\|_\mathcal{G}\end{aligned}$$
on the event $\{X \neq 0\}$. Thus $\|Y - W\|_\mathcal{G} \le \|Y - U\|_\mathcal{G}$ for any $U \in L^\infty(\mathcal{G})$ and hence $W \in \mathbb{E}[Y \,|\, \mathcal{G}]$.

We now show the converse inclusion $X\mathbb{E}[Y \,|\, \mathcal{G}] \subseteq \mathbb{E}[XY \,|\, \mathcal{G}]$.

Choose $W \in \mathbb{E}[Y \,|\, \mathcal{G}]$. We need to show that $XW \in \mathbb{E}[XY \,|\, \mathcal{G}]$. Consider $U \in L^\infty(\mathcal{G})$. Put $V = (U/X)\,\mathbf{1}\{X \neq 0\}$. We have $\|Y -$

$W\|_{\mathcal{G}} \le \|Y - V\|_{\mathcal{G}}$ by the assumption that $W \in \mathbb{E}[Y \mid \mathcal{G}]$. From Lemma 4.3(i)+(ii),

$$\begin{aligned}\|XY - XW\|_{\mathcal{G}} &= |X|\|Y - W\|_{\mathcal{G}} \le |X|\|Y - V\|_{\mathcal{G}} = \|XY - XV\|_{\mathcal{G}} \\ &= \|XY - U\|_{\mathcal{G}}\,\mathbf{1}\{X \neq 0\} \le \|XY - U\|_{\mathcal{G}},\end{aligned}$$

as required.

The proof of the claim $\mathbb{E}[X + Y \mid \mathcal{G}] = X + \mathbb{E}[Y \mid \mathcal{G}]$ is similar but easier, so we omit it.

Parts (ii) and (iii) follow straightforwardly from parts (ii) and (iii) of Lemma 4.3. □

Proposition 6.2. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Suppose that $X \in L^\infty$ is independent of $\mathcal{G}$. Then $\mathbb{E}[X \mid \mathcal{G}]$ is the set of random variables $Y \in L^\infty(\mathcal{G})$ that take values in $\mathbb{E}[X]$.*

Proof. Observe for any $Z \in L^\infty(\mathcal{G})$, that, by the assumption of independence of X from $\mathcal{G}$,

$$\begin{aligned}\|X - Z\|_{\mathcal{G}}(\omega) &= \sup_p \left(\mathbb{E}[|X - Z|^p \mid \mathcal{G}](\omega)\right)^{\frac{1}{p}} \\ &= \sup_p \left(\int |x - Z(\omega)|^p \,\mathbb{P}\{X \in dx\}\right)^{\frac{1}{p}} \\ &= \sup\{|x - Z(\omega)| : x \in \mathrm{supp} X\} \\ &\begin{cases} = \varepsilon(X), & \text{if } Z(\omega) \in \mathbb{E}[X], \\ > \varepsilon(X), & \text{otherwise,}\end{cases}\end{aligned}$$

and the result follows. □

7. Conditional spread and the tower property

Definition 7.1. Given $X \in L^\infty$ and a sub-σ-field $\mathcal{G}$ of $\mathcal{F}$, let $\varepsilon(X, \mathcal{G})$ denote the common value of $\|X - Y\|_{\mathcal{G}}$ for $Y \in \mathbb{E}[X \mid \mathcal{G}]$.

Lemma 7.2. *If $X \in L^\infty$ and a $\mathcal{G} \subseteq \mathcal{H}$ are sub-σ-fields of $\mathcal{F}$, then $\varepsilon(X, \mathcal{H}) \le \varepsilon(X, \mathcal{G})$.*

Proof. Suppose that $V \in \mathbb{E}[X \mid \mathcal{G}]$ and $W \in \mathbb{E}[X \mid \mathcal{H}]$. From Lemma 4.4,

$$\varepsilon(X, \mathcal{H}) = \|X - W\|_{\mathcal{H}} \le \|X - V\|_{\mathcal{H}} \le \|X - V\|_{\mathcal{G}} = \varepsilon(X, \mathcal{G}).$$

□

Lemma 7.3. *A random variable Y belongs to $\mathbb{E}[X \mid \mathcal{G}]$ if and only if $Y \in L^\infty(\mathcal{G})$ and $|X - Y| \le \varepsilon(X, \mathcal{G})$.*

Proof. Suppose Y is in $\mathbb{E}[X \mid \mathcal{G}]$. By definition, $Y \in L^\infty(\mathcal{G})$. By Lemma Lemma 4.4, $|X - Y| = \|X - Y\|_{\mathcal{F}} \le \|X - Y\|_{\mathcal{G}} = \varepsilon(X, \mathcal{G})$.

The converse is immediate from Lemma 3.2(ii). □

Lemma 7.4. *Suppose that $X \in L^\infty$, $\mathcal{G} \subseteq \mathcal{H}$ are sub-σ-fields of $\mathcal{F}$, and $Y \in \mathbb{E}[X \mid \mathcal{H}]$. Then $\varepsilon(Y, \mathcal{G}) \le \varepsilon(X, \mathcal{G})$.*

Proof. Consider $Z \in \mathbb{E}[X \mid \mathcal{G}]$. By Lemma 7.3 and Lemma 7.2

$$|Y - Z| \le |X - Y| \vee |X - Z| \le \varepsilon(X, \mathcal{H}) \vee \varepsilon(X, \mathcal{G}) = \varepsilon(X, \mathcal{G}).$$

By Lemma 3.2(ii), $\varepsilon(Y, \mathcal{G}) \le \|Y - Z\|_{\mathcal{G}} \le \varepsilon(X, \mathcal{G})$. □

Theorem 7.5. *Suppose that $X \in L^\infty$ and $\mathcal{G} \subseteq \mathcal{H}$ are sub-σ-fields of $\mathcal{F}$. If $Y \in \mathbb{E}[X \mid \mathcal{H}]$ and $Z \in \mathbb{E}[Y \mid \mathcal{G}]$, then $Z \in \mathbb{E}[X \mid \mathcal{G}]$.*

Proof. By Lemma 7.3, Lemma 7.4, and Lemma 7.2,

$$|X - Z| \le |X - Y| \vee |Y - Z| \le \varepsilon(X, \mathcal{H}) \vee \varepsilon(Y, \mathcal{G}) \le \varepsilon(X, \mathcal{G}).$$

Thus Z is in $\mathbb{E}[X \mid \mathcal{G}]$ by another application of Lemma 7.3. □

8. Continuity of conditional expectation

Definition 8.1. Define the Hausdorff distance between two subsets A and B of L^∞ to be

$$D_H(A, B) := \sup_{X \in A} \inf_{Y \in B} \|X - Y\|_\infty \vee \sup_{Y \in B} \inf_{X \in A} \|Y - X\|_\infty.$$

Lemma 8.2. *Suppose that A, B, C are subsets of L^∞. Then*

$$D_H(A + C, B + C) \le D_H(A, B).$$

Proof. Suppose that $D_H(A, B) < \delta$ for some $\delta \ge 0$. By definition, for every $X \in A$ there is a $Y \in B$ with $\|X - Y\|_\infty < \delta$, and similarly with the roles of A and B reversed. If $U \in A + C$, then $U = X + W$ for some $X \in A$ and $W \in C$. We know there is $Y \in B$ such that $\|X - Y\|_\infty < \delta$. Note that $V := Y + W \in B + C$ and $\|U - V\|_\infty = \|X - Y\|_\infty < \delta$. A similar argument with the roles of A and B reversed shows that $D_H(A + C, B + C) < \delta$. □

Theorem 8.3. *Suppose that $X, Y \in L^\infty$ and $\mathcal{G}$ is a sub-σ-field of $\mathcal{F}$. Then $D_H(\mathbb{E}[X \mid \mathcal{G}], \mathbb{E}[Y \mid \mathcal{G}]) \le \|X - Y\|_\infty$.*

Proof. Choose $U \in \mathbb{E}[X \mid \mathcal{G}]$ and $V \in \mathbb{E}[Y \mid \mathcal{G}]$. From Lemma 4.3(iv),

$$\varepsilon(Y, \mathcal{G}) \le \|Y - U\|_{\mathcal{G}} \le \|X - U\|_{\mathcal{G}} \vee \|X - Y\|_{\mathcal{G}} = \varepsilon(X, \mathcal{G}) \vee \|X - Y\|_{\mathcal{G}}$$

and

$$\varepsilon(X, \mathcal{G}) \le \|X - V\|_{\mathcal{G}} \le \|Y - V\|_{\mathcal{G}} \vee \|X - Y\|_{\mathcal{G}} = \varepsilon(Y, \mathcal{G}) \vee \|X - Y\|_{\mathcal{G}}.$$

It follows that $\varepsilon(X,\mathcal{G}) = \varepsilon(Y,\mathcal{G})$ on the event $M := \{\|X-Y\|_{\mathcal{G}} < \varepsilon(X,\mathcal{G}) \vee \varepsilon(Y,\mathcal{G})\}$ and

$$\varepsilon(X,\mathcal{G}) = \|Y-U\|_{\mathcal{G}} = \|X-U\|_{\mathcal{G}} = \varepsilon(X,\mathcal{G})$$

and

$$\varepsilon(Y,\mathcal{G}) = \|X-V\|_{\mathcal{G}} = \|Y-V\|_{\mathcal{G}} = \varepsilon(Y,\mathcal{G})$$

on M.

By Proposition 6.1, $U\,\mathbf{1}_M \in \mathbb{E}[Y\,\mathbf{1}_M \,|\, \mathcal{G}] = \mathbf{1}_M\mathbb{E}[Y \,|\, \mathcal{G}]$ and $V\,\mathbf{1}_M \in \mathbb{E}[X\,\mathbf{1}_M \,|\, \mathcal{G}] = \mathbf{1}_M\mathbb{E}[X \,|\, \mathcal{G}]$. Thus $\mathbf{1}_M\mathbb{E}[X \,|\, \mathcal{G}] = \mathbf{1}_M\mathbb{E}[Y \,|\, \mathcal{G}]$.

Furthermore, on the event $N := \{\|X-Y\|_{\mathcal{G}} \geq \varepsilon(X,\mathcal{G}) \vee \varepsilon(Y,\mathcal{G})\}$

$$\begin{aligned}\|U-V\|_\infty &\leq \|U-X\|_\infty \vee \|X-Y\|_\infty \vee \|Y-V\|_\infty \\ &\leq \varepsilon(X,\mathcal{G}) \vee \|X-Y\|_\infty \vee \varepsilon(Y,\mathcal{G}) \\ &\leq \|X-Y\|_\infty,\end{aligned}$$

and so $\|U\,\mathbf{1}_N - V\,\mathbf{1}_N\|_\infty \leq \|X\,\mathbf{1}_N - Y\,\mathbf{1}_N\|_\infty \leq \|X-Y\|_\infty$. Therefore,

$$D_H(\mathbf{1}_N\mathbb{E}[X \,|\, \mathcal{G}], \mathbf{1}_N\mathbb{E}[Y \,|\, \mathcal{G}]) \leq \|X-Y\|_\infty.$$

By Proposition 6.1(iii), $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbf{1}_M\mathbb{E}[X \,|\, \mathcal{G}] + \mathbf{1}_N\mathbb{E}[X \,|\, \mathcal{G}]$, and similarly for Y. The result now follows from Lemma 8.2. □

9. Martingales

Definition 9.1. Let $\{\mathcal{F}_n\}_{n=0}^\infty$ be a filtration of sub-σ-fields of $\mathcal{F}$. A sequence of random variables $\{X_n\}_{n=0}^\infty$ is a *martingale* if there exists $X \in L^\infty$ such that $X_n \in \mathbb{E}[X \,|\, \mathcal{F}_n]$ for all n (in particular, $X_n \in L^\infty(\mathcal{F}_n)$).

Remark 9.2. Note that our definition does not imply that $X_n \in \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n]$ for all n. For example, suppose that $\mathcal{F}_n := \{\emptyset, \Omega\}$ for all n but X is not almost surely constant, then we obtain a martingale by taking X_n to be any constant in the ball $\mathbb{E}[X]$, but we only have $X_n \in \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n]$ for all n if $X_0 = X_1 = X_2 = \ldots$.

Many of the usual real-valued examples of martingales have $\mathbb{K}$-valued counterparts.

Example 9.3. Let $\{Y_n\}_{n=0}^\infty$ be a sequence of independent random variables in L^∞ with $0_{\mathbb{K}} \in \mathbb{E}[Y_n]$ for all n. Suppose that $\sum_{k=0}^\infty Y_k$ converges in L^∞ (by the strong triangle inequality and the completeness of L^∞, this is equivalent to $\lim_{n\to\infty} \|Y_n\|_\infty = 0$). Set $\mathcal{F}_n := \sigma\{Y_0, Y_1, \ldots, Y_n\}$. Put $X_n := \sum_{k=0}^n Y_k$ and $X_n := \sum_{k=0}^\infty Y_k$ It follows from the second claim of Proposition 6.1(i) that $X_n \in \mathbb{E}[X \,|\, \mathcal{F}_n]$ for all n and hence $\{X_n\}_{n=0}^\infty$ is a martingale.

Example 9.4. Let $\{Y_n\}_{n=0}^{\infty}$ be a sequence of independent random variables in L^∞ with $1_{\mathbb{K}} \in \mathbb{E}[Y_n]$ for all n. Suppose that $\prod_{k=0}^{\infty} Y_k$ converges in L^∞ (by the strong triangle inequality and the completeness of L^∞, this is equivalent to $\lim_{n\to\infty} \|Y_n - 1_{\mathbb{K}}\|_\infty = 0$). Set $\mathcal{F}_n := \sigma\{Y_0, Y_1, \ldots, Y_n\}$. Put $X_n := \prod_{k=0}^{n} Y_k$ and $X := \prod_{k=0}^{\infty} Y_k$. It follows from the first claim of Proposition 6.1(i) that $X_n \in \mathbb{E}[X \mid \mathcal{F}_n]$ for all n and hence $\{X_n\}_{n=0}^{\infty}$ is a martingale.

Example 9.5. Let $\{Z_n\}_{n=0}^{\infty}$ be a discrete time Markov chain with countable state space E and transition matrix P. Set $\mathcal{F}_n := \sigma\{Z_0, Z_1, \ldots, Z_n\}$. Say that $f : E \to \mathbb{K}$ is *harmonic* if f is bounded and for all $i \in E$ the expectation of f with respect to the probability measure $P(i, \cdot)$ contains $f(i)$ (that is, if $f(i)$ is belongs to the smallest ball containing the set $\{f(j) : P(i,j) > 0\}$). Fix $N \in \{0, 1, 2, \ldots\}$. Then $\{X_n\}_{n=0}^{\infty} := \{f(Z_{n\wedge N})\}_{n=0}^{\infty}$ is a martingale.

10. Optional sampling theorem

Theorem 10.1. *Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a filtration. Suppose that $X \in L^\infty$ and $\{X_n\}_{n=0}^{\infty}$ is a martingale with $X_n \in \mathbb{E}[X \mid \mathcal{F}_n]$ for all n. If T is a stopping time, then $X_T \in \mathbb{E}[X \mid \mathcal{F}_T]$.*

Proof. It follows from Lemma 4.5 that $\mathbf{1}\{T = n\}\mathbb{E}[X \mid \mathcal{F}_T] = \mathbf{1}\{T = n\}\mathbb{E}[X \mid \mathcal{F}_n]$ and hence, by Proposition 6.1(iii),

$$\begin{aligned}
\mathbb{E}[X \mid \mathcal{F}_T] &= \mathbb{E}\left[\sum_n X\,\mathbf{1}\{T = n\} \mid \mathcal{F}_T\right] \\
&= \sum_n \mathbf{1}\{T = n\}\mathbb{E}[X \mid \mathcal{F}_T] \\
&= \sum_n \mathbf{1}\{T = n\}\mathbb{E}[X \mid \mathcal{F}_n] \\
&\ni \sum_n \mathbf{1}\{T = n\} X_n \\
&= X_T.
\end{aligned}$$

□

11. Martingale convergence

Theorem 11.1. *Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be a filtration. Suppose that $X \in L^\infty$ and $\{X_n\}_{n=0}^{\infty}$ is a martingale with $X_n \in \mathbb{E}[X \mid \mathcal{F}_n]$ for all n. If X is in the closure of $\bigcup_{n=1}^{\infty} L^\infty(\mathcal{F}_n)$, then $\lim_{n\to\infty} \|X_n - X\|_\infty = 0$ (in particular, $\{X_n\}_{n=0}^{\infty}$ converges to X almost surely).*

Proof. Since X is in the closure of $\bigcup_{n=1}^{\infty} L^\infty(\mathcal{F}_n)$, for each $\varepsilon > 0$ there exists $Y \in L^\infty(\mathcal{F}_N)$ for some N such that $\|X - Y\|_\infty < \varepsilon$. Because $\mathcal{F}_N \subseteq \mathcal{F}_n$ for $n > N$, $Y \in L^\infty(\mathcal{F}_n)$ for $n \geq N$.

By Theorem 8.3, $D_H(E[X \mid \mathcal{F}_n], E[Y \mid \mathcal{F}_n]) < \varepsilon$ for $n \geq N$. However, $E[Y \mid \mathcal{F}_n]$ consists of the single point Y, and so the Hausdorff distance is simply $\sup\{\|W - Y\|_\infty : W \in E[X \mid \mathcal{F}_n]\}$. Thus

$$\|X_n - X\|_\infty \leq \|X_n - Y\|_\infty \vee \|Y - X\|_\infty < \varepsilon$$

for $n \geq N$. □

References

[Eva89] Steven N. Evans, *Local field Gaussian measures*, Seminar on Stochastic Processes, 1988 (Gainesville, FL, 1988), Progr. Probab., vol. 17, Birkhäuser Boston, Boston, MA, 1989, pp. 121–160. MR **MR990478 (91e:**60121)

[Eva91] ———, *Equivalence and perpendicularity of local field Gaussian measures*, Seminar on Stochastic Processes, 1990 (Vancouver, BC, 1990), Progr. Probab., vol. 24, Birkhäuser Boston, Boston, MA, 1991, pp. 173–181. MR **MR1118442 (93m:**60078)

[Eva93] ———, *Local field Brownian motion*, J. Theoret. Probab. **6** (1993), no. 4, 817–850. MR **MR1245397 (94i:**60098)

[Eva95] ———, *p-adic white noise, chaos expansions, and stochastic integration*, Probability measures on groups and related structures, XI (Oberwolfach, 1994), World Sci. Publishing, River Edge, NJ, 1995, pp. 102–115. MR **MR1414928 (98e:**60083)

[Eva01a] ———, *Local field U-statistics*, Algebraic methods in statistics and probability (Notre Dame, IN, 2000), Contemp. Math., vol. 287, Amer. Math. Soc., Providence, RI, 2001, pp. 75–81. MR **MR1873668 (2003b:**60014)

[Eva01b] ———, *Local fields, Gaussian measures, and Brownian motions*, Topics in probability and Lie groups: boundary theory, CRM Proc. Lecture Notes, vol. 28, Amer. Math. Soc., Providence, RI, 2001, pp. 11–50. MR **MR1832433 (2003e:**60012)

[Eva02] ———, *Elementary divisors and determinants of random matrices over a local field*, Stochastic Process. Appl. **102** (2002), no. 1, 89–102. MR **MR1934156 (2004c:**15041)

[Eva06] Steven N. Evans, *The expected number of zeros of a random system of p-adic polynomials*, U.C. Berkeley Department of Statistics Technical Report No. 699 and American Institute of Mathematics preprint number AIM 2006-9, 2006.

[Khr97] Andrei Khrennikov, *Non-Archimedean analysis: quantum paradoxes, dynamical systems and biological models*, Mathematics and its Applications, vol. 427, Kluwer Academic Publishers, Dordrecht, 1997. MR **MR1746953 (2001h:**81004)

[KN04] Andrei Yu. Khrennikov and Marcus Nilson, *p-adic deterministic and random dynamics*, Mathematics and its Applications, vol. 574, Kluwer Academic Publishers, Dordrecht, 2004. MR **MR2105195 (2005h:**37102)

[Koc01] Anatoly N. Kochubei, *Pseudo-differential equations and stochastics over non-Archimedean fields*, Monographs and Textbooks in Pure and Applied Mathematics, vol. 244, Marcel Dekker Inc., New York, 2001. MR **MR1848777 (2003b:**35220)

[Nev75] J. Neveu, *Discrete-parameter martingales*, revised ed., North-Holland Publishing Co., Amsterdam, 1975, Translated from the French by T. P. Speed, North-Holland Mathematical Library, Vol. 10. MR MR0402915 (53 #6729)

[Sch84] W. H. Schikhof, *Ultrametric calculus*, Cambridge Studies in Advanced Mathematics, vol. 4, Cambridge University Press, Cambridge, 1984, An introduction to p-adic analysis. MR **MR791759 (86j:**11104)

[Tai75] M. H. Taibleson, *Fourier analysis on local fields*, Princeton University Press, Princeton, N.J., 1975. MR MR0487295 (58 #6943)

[vR78] A. C. M. van Rooij, *Non-Archimedean functional analysis*, Monographs and Textbooks in Pure and Applied Math., vol. 51, Marcel Dekker Inc., New York, 1978. MR **MR512894 (81a:**46084)

Department of Statistics #3860, University of California at Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, U.S.A.

E-mail address: evans@stat.berkeley.edu

URL: http://www.stat.berkeley.edu/users/evans/

E-mail address: tlid@berkeley.edu