Now that we have shown that every Boolean algebra can be extended to a complete one equipped with a countably additive probability, let us abuse notation and assume from now on that $\mathfrak{B}$ is complete and $\mathbb{P}$ countably additive on it, with all previous operations defined, along with $\mathbf{0} = \varnothing$ and $\mathbf{1}$. Our next goal would be to define random variables and integrations (expected values) on them. Without directly working with points, we naturally consider "random variables" characterized by their values on various regions. To this end, we say a collection of elements $\{x_i\}_{i\in I}$, countable or finite, **partitions** $\mathfrak{B}$ if they are pairwise disjoint, i.e., $x_i \wedge x_j = \varnothing$ for $i \neq j$, and $\bigvee_{i\in I} x_i = \mathbf{1}$. It follows by countable additivity that any partition $\{x_i\}_{i\in I}$ satisfies $\sum_{i\in I} \mathbb{P}(x_i) = 1$. The following definitions are straightforward:

- A **simple random variable** (resp. **elementary random variable**) is a real-valued function X defined on a finite (resp. countable) partition $\{x_n\}$ of $\mathfrak{B}$ that maps each x_i to a constant value. We denote the collection of elementary random variables by $\mathfrak{E}$.
- A **constant random variable** is a simple random variable defined on $\{\mathbf{1}, \varnothing\}$.
- Given $x \in \mathfrak{B}$, a corresponding **indicator random variable** I_x can be defined as the simple random variable on $\{x, x^c\}$ with $I_x(x) = 1$ and $I_x(x^c) = 0$.

For simplicity, unless otherwise stated, when given X and is corresponding partition $\{x_n\}$, we assume X does not take duplicate values on different regions: if $x_i \neq x_j$ then $X(x_j) \neq X(x_j)$, for otherwise we can always take out x_i, x_j from $\{x_n\}$ and insert $x_i \vee x_j$ into it.

In some sense, we are defining elementary random variables based on their indicator decomposition — we will justify this statement later. For now, we compare two random variables X on $\{x_n\}$ and Y on $\{y_n\}$ by considering the following "finer" partition consisting of all non-zero meets of $x_i \wedge y_j$. Let us write this as $\{z_m\} \coloneqq \{x_n\} \oplus \{y_n\}$. It is easy to see that $\{z_m\}$ is still a countable partition of $\mathfrak{B}$. Then both X and Y can be defined with respect to $\{z_m\}$ (though this violates our just-stated no duplicate assumption), and we define their sum to be the elementary random variable

$$X \leqslant Y \text{ if } X(z_i) \leqslant Y(z_i) \text{ for each } z_i \in \{z_m\} = \{x_n\} \oplus \{y_n\}.$$

We may define $\geqslant$ analogously. Elementary arithmetic of X and Y can be defined via $\{z_m\}$ as well, and the idea behind each should be self-explanatory:

- Addition: $X + Y$ is defined to be the elementary random variable on $\{z_m\}$ by $(X+Y)(z_i) \coloneqq X(z_i) + Y(z_i)$, or equivalently, $(X+Y)(x_i \wedge y_j) \coloneqq X(x_i) + Y(y_j)$.
- Scalar multiplication: given $c \in \mathbb{R}$, define the scalar multiple cX as $(cX)(x_i) \coloneqq cX(x_i)$ for each i.
- Multiplication: XY is defined to be the elementary random variable on $\{z_m\}$ by $(XY)(z_i) \coloneqq X(z_i)Y(z_i)$, or equivalently, $(X+Y)(x_i \wedge y_j) \coloneqq X(x_i)Y(y_j)$.

Many nice properties of elementary random variables follow from these definitions: symmetry, linearity, distributivity, and so on. We define the "region-wise" minimum, $X \wedge Y$ and $X \vee Y$, by

$$(X \wedge Y)(z_i) = \min(X(z_i), Y(z_i))$$

or equivalently $(X \wedge Y)(x_i \wedge y_j) = \min(X(x_i), Y(y_j))$, and likewise a "region-wise" maximum $X \vee Y$. Next, we define the positive and negative parts of X, denoted X^+, X^-, by

$$X^+ \coloneqq X \vee 0 \qquad \text{and} \qquad X^- = -(X \wedge 0)$$

where 0 here represents the constant random variable taking value 0. It follows that $X = X^+ - X^-$ and we define the absolute value of X to be $|X| = X^+ \vee X^-$.

So far, we have analyzed simple random variables on $(\mathfrak{B}, \mathbb{P})$ and have shown that they behave much like simple random variables defined in a pointy setting, e.g., a probability space defined on Borel sets in $\mathbb{R}^n$. With finite addition of simple random variables defined, given X defined on $\{x_i\}_{i=1}^n$, we can naturally represent X as a linear combination of indicator random variables $X = \sum_{i=1}^n X(x_i) I_{x_i}$. Clearly, we would want a similar expression for elementary random variables, where the finite sum is replaced by an infinite sum. To do so, we need to justify the limit and before that, define what it means for a sequence of (elementary) random variables to converge.
We say a sequence of elementary random variables $\{X_n\} \subset \mathfrak{E}$ **converges** to X, written $X_n \to X$, if there exists a decreasing sequence $\{Y_n\} \subset \mathfrak{E}$ such that

$$\bigwedge_{k \geqslant 1} Y_k \text{ exists and equals } 0 \qquad \text{and} \qquad |X_n - X| \leqslant Y_n.$$

Intuitively, this is a relatively weak sense of convergence — to draw some analogy to the pointy spaces, this notion is somewhat similar to pointwise convergence: $\bigwedge_{k \geqslant 1} Y_k$ is the infinite term-wise minimum of the sequence $\{Y_n\}$, so if that equals 0, then $Y_n \to 0$ pointwise. It is however slightly stronger than pointwise convergence, in the sense that it still imposes some kind of uniform bound. In pointy probability theory a famous example showing almost sure convergence does not imply convergence in expectation is $X_n = n \cdot I_{[0,1/n]}$, where because of the unboundedeess of n, a point mass "escapes" at 0. Here, such things cannot happen because we required $\{Y_n\}$ to decrease a priori.

Analogously, we define $\{X_n\} \subset \mathfrak{E}$ to be **Cauchy** if there exists a decreasing sequence $\{Y_n\} \subset \mathfrak{E}$ such that for each n and all $i, j > n$, $|X_i, X_j| \leqslant Y_n$.

Returning to our main goal, let $X \in \mathfrak{E}$ be defined on a countable $\{x_n\}$. To show that X can be represented by $\sum_{i=1}^\infty X(x_i) I_{x_i}$, we approximate this infinite sum by finite sums $\sum_{i=1}^n X(x_i) I_{x_i}$. We define the partial joins $y_n = \bigvee_{i=1}^n x_i$ and notice that the partial sums $\sum_{i=1}^n X(x_i) I_{x_i} = I_{y_n} X$. We apply this convergence to I_{y_n} and show it converges to $\mathbf{1}$, the constant random variable taking value 1 (or equivalently the indicator of $\mathbf{1} \in \mathfrak{B}$):

$$|I_{y_n} - \mathbf{1}| = \bigwedge_{i=1}^n x_i^c \qquad \text{and} \qquad \bigwedge_{i=1}^n x_i^c = \varnothing$$

since the infinite join does not contain any x_i, but the x_i's make up the entire space. Therefore $I_{y_n} \to \mathbf{1}$. By the same token we see that $I_{y_n} X \to X$, so the proof is complete, and we claim that

$$X = \sum_{k \geqslant 1} X(x_k) I_{x_k}$$

is indeed an indicator representation of X.

Our next question is, how do we define a more general form of random variables using approximations of elementary random variables? In a pointy setting, given $X \geqslant 0$, it is well-known that X can be approximated pointwise by a sequence $\{X_n\}$ of monotone increasing simple random variables by considering $X\mathbf{1}[X \leqslant n]$ and rounding the values of this variable down modulo 2^{-n}, so that the resulting random variable only takes values of multiples of 2^{-n}, with minimum 0 and maximum n.
In our setting, we also consider convergence of elementary random variables, but instead of appealing to dyadic numbers we again consider equivalence classes of convergent elementary random variables based on the "limit," as we once did during the process of Cauchy completing our initial Boolean algebra.

To achieve, this, we first define sequence-wise operations of sequences of elementary random variables. Abusing the notations, given $\{X_n\}, \{Y_n\} \subset \mathfrak{E}$, we define addition, multiplication, scalar multiplication, join / pointwise maximum, and meet / pointwise minimum element-wise, i.e., $\{X_n\} + \{Y_n\} := \{X_n + Y_n\}$ and so on. Once again, we consider the quotient space of Cauchy sequences modulo sequences that converge to zero. Notation-wise, we define $\mathfrak{C}(\mathfrak{E})$ to be the space of Cauchy sequences of elementary random variables, with operations defined component-wise, and $\mathfrak{C}_0(\mathfrak{E})$ the space of sequences of elementary random variables converging to 0 (the zero constant variable). Then we define the space of **general random variables** (or just **random variable**) to be $\mathfrak{X} := \mathfrak{C}(\mathfrak{E})/\mathfrak{C}_0(\mathfrak{E})$. In other words, we identify each general random variable with the collection of all Cauchy sequences whose "limit" agree, and we shall a general random variable by an equivalence class written as $X = [\{X_n\}]$. This way, $\mathfrak{X}$ has many "nice" properties that align with our intuition. Furthermore, when defining $\mathfrak{X}$ we used a notion of convergence akin to pointwise convergence, but Kappos (1970) has shown that if X_n converges to X, then there exists another sequence $\{Y_n\}$ that converges to X **uniformly**. That is, there exist constant random variables $\{c_n\}$, $c_n \downarrow 0$, such that $|X_n - X| \leqslant c_n$. In the following sections we consider several properties of random variables defined in this manner, and show that they align nicely with our intuition and pointy counterparts.

We first explore the notion of distribution. In a pointy setting it is well-known that a random variable X is characterized by its distribution function $F(x) = \mathbb{P}(X < x)$. Here, we will show that a similar notions holds on both $\mathfrak{E}$ and $\mathfrak{X}$. Firstly, for any elementary random variable $X \in \mathfrak{E}$ defined on $\{x_i\}$and $c \in \mathbb{R}$, we define

$$\overline{D}_X(c) := [X \leqslant c] = \bigvee_{X(x_i) \leqslant c} x_i \qquad \text{and} \qquad \underline{D}_X(c) := [X < c] = \bigvee_{X(x_i) < c} x_i.$$

Both $\overline{D}_X$ and $\underline{D}_X$ are monotone increasing functions, in the sense that if $c_1 < c_2$ then $\bigvee_{X(x_i) \leqslant c_1} x_i$ is contained in $\bigvee_{X(x_i) \leqslant c_2} x_i$. Since $\{x_i\}$ can be labeled arbitrarily, we may WLOG assume that $X(x_i) < X(x_j)$ for $i < j$, so that

$$\bigvee_{X(x_i) \leqslant c} x_i = \bigvee_{i=1}^{k(c)} x_i$$

where $k(c)$ is the largest index, 0 if nonexistent, such that $X(x_i) \leqslant c$, and likewise for strict inequality and $\underline{D}_X$. Clearly $\lim_{c \to -\infty} k(c) = 0$ and $\lim_{c \to \infty} k(c) = \infty$. But then

$$\lim_{c \to -\infty} \overline{D}_X(c) = \lim_{c \to -\infty} \underline{D}_X(c) = \lim_{k(c) \to 0} \bigvee_{i=1}^{k(c)} x_i = \varnothing \qquad \text{and} \qquad \lim_{c \to \infty} \overline{D}_X(c) = \lim_{c \to \infty} \underline{D}_X(c) = \mathbf{1}.$$

What happens if we now consider $\{X_n\} \subset \mathfrak{E}$ and the corresponding $X = [\{X_n\}] \in \mathfrak{X}$? We can extend the definition of $\overline{D}, \underline{D}$ by defining

$$\overline{D}_X(c) := \limsup_{n \to \infty} \overline{D}_{X_n}(c) \qquad \text{and} \qquad \underline{D}_X(c) := \liminf_{n \to \infty} \underline{D}_{X_n}(c),$$

where the $\limsup, \liminf$ are defined with nested operations of $\vee$ and $\wedge$. This definition is compatible with the version defined on $\mathfrak{E}$ because for any $X \in \mathfrak{E}$ we can simply consider the constant sequence $\{X, X, \cdots\}$. To show that this definition is well-defined, we consider $\{X_n\} \in \mathfrak{C}(\mathfrak{E})$ Cauchy and $\{Y_n\} \in \mathfrak{C}_0(\mathfrak{E})$, a sequence converging to 0. Since we may replace $\{Y_n\}$ with a sequence uniformly converging to 0, the difference between X_n and $X_n + Y_n$ is uniformly bounded, and so are $\overline{D}_{X_n}$ and $\overline{D}_{X_n + Y_n}$ (and likewise for $\underline{D}$). Letting $\epsilon \downarrow 0$ the claim follows.
Therefore, we are able to characterize any $X \in \mathfrak{X}$ by its **distribution function** $F : \mathbb{R} \to \mathfrak{B}$ defined by $F(c) = [X < c]$. This is a monotone function with limits $F(-\infty) = \lim_{c \to -\infty} F(c) = \varnothing$ and $F(\infty) = \mathbf{1}$. We choose $[X < c]$ over $[X \leqslant c]$ because the former is right continuous, i.e., $\lim_{x \downarrow c} F(x) = F(c)$.

An important result that we can derive via distribution functions is:

Theorem

$\mathfrak{X}$ is complete with respect to arbitrary joins and meets taken over a collection of uniformly bounded random variables. In other words, if $\{X_i\}_{i\in\mathfrak{I}} \subset \mathfrak{X}$ is indexed over any arbitrary I, and there exists $M \geqslant 0 \in \mathfrak{X}$ such that $-M \leqslant X_i \leqslant M$ for all i, then both $\bigvee_{i\in I} X_i$ and $\bigwedge_{i\in I} X_i$ exist and are in $\mathfrak{X}$.

Proof. $\mathfrak{B}$ is complete, so the arbitrary join $\bigwedge_{i\in I} \underline{D}_{X_i}(c)$ exists for each c. Define $\underline{D}_X(c)$ to be the right limits of $\bigwedge_{i\in I} \underline{D}_{X_i}(x)$, i.e.,

$$\underline{D}_X(c) := \lim_{x\downarrow c} \bigwedge_{i\in I} \underline{D}_{X_i}(x).$$

Then $\underline{D}$ satisfies all the criterion for a distribution function. And it follows by completeness of $\mathfrak{B}$ that the random variable corresponding to $\underline{D}_X$ must coincide with $\bigvee_{i\in I} X_i$, so the closure with respect to arbitrary join is proven. The other case $\bigwedge_{i\in I} X_i$ is analogous. □

Note that the assumption of boundedness is necessary. Since if $X \leqslant Y$, $\underline{D}_X \geqslant \underline{D}_Y$, $-M \leqslant X_i \leqslant M$ implies that for all i, $\underline{D}_M \leqslant \underline{D}_{X_i} \leqslant \underline{D}_{-M}$. Without this assumption the claim fails, for we may construct an example where $\underline{D}_X$ becomes uniformly **0**, breaking $\lim_{c\to\infty} \underline{D}_X(c) = \mathbf{1}$.

Let us now turn our attention to defining expected values of random variables. As usual, we start with elementary ones. Let $X \in \mathfrak{E}$ and write X as $\sum_{k\geqslant 1} X(x_k) I_{x_k}$. If $\sum_{k\geqslant 1} |X(x_k)|\mathbb{P}(x_k) < \infty$, then we define the **expected value** of X, written $\mathbb{E}X$, to be $\sum_{k\geqslant 1} X(x_k)\mathbb{P}(x_k)$. Note that not all $X \in \mathfrak{E}$ possess expected values, since not all of them have absolutely convergent indicator representations. It is clear that the collection of elementary variables with well-defined expected values is closed under addition, scalar multiplication, and maximum ($\vee$) and minimum ($\wedge$). More generally, given $X \in \mathfrak{X}$, we know that there exists a sequence $\{X_n\} \subset \mathfrak{E}$ converging uniformly to X. We say X possesses an expected value if in addition each X_n possesses an expected value. Intuitively, uniform convergence preserves limit of expected values, and this is indeed true: uniform convergence, along with triangle inequality, implies

$$|\mathbb{E}X_n - \mathbb{E}X| \leqslant \mathbb{E}|X_n - X| \to 0.$$

This is a natural extension of the expected value defined on $\mathfrak{E}$, and therefore it inherits the algebraic structure. In particular, the previously mentioned algebraic operators are also well-defined on $\mathfrak{X}$ with respect to expected values. We define L^1 to be the space of all random variables X with $\mathbb{E}|X| < \infty$. Clearly L^1 is also closed under addition, scalar multiplication, maximum, and minimum.
Our next goal is to recover the famous Convergence Theorems.

Theorem: Dominated Convergence Theorem

If $X_n \geqslant 0, X_n \in L^1$ is monotone decreasing with $X_n \to 0$, then the expectations converge as well: $\mathbb{E}X_n \downarrow 0$.

Proof. We will prove this claim via a multi-step procedure: first we show it holds for simple random variables, then elementary random variables, and finally, (general) random variables.

STEP 1: SIMPLE RANDOM VARIABLES. Let each X_n be simple, defined on $\{x_{(n,i)}\}_{i=1}^{c(n)}$. Using indicator representation, we write $X_n = \sum_{i=1}^{c(n)} X_n(x_{(n,i)}) I_{x_{(n,i)}}$. Further WLOG assume that for each fixed n, the $x_{(n,i)}$'s are arranged in the decreasing order based on $X_n(x_{(n,i)})$, i.e., $X_n(x_{(n,i)}) \geqslant X_n(x_{(n,i+1)})$. Let $\epsilon > 0$ be given. Our goal is to show that for sufficiently large n, $\mathbb{E}X_n < \epsilon$. The idea behind the proof is that we show as $n \to \infty$, X_n takes value $> \epsilon$ on a sufficiently small region, whose contribution to $\mathbb{E}X_n$ can be controlled, whereas X_n is sufficiently small on the remainder of the space, and consequently its contribution to $\mathbb{E}X_n$ is also controllable.

Let $M = X_1(x_{1,1})$. By monotonicity of both $\{x_{n,1}, x_{n,2}, \cdots\}$ and $\{X_n\}$, we know $M \geqslant \mathbb{E}X_1 \geqslant \mathbb{E}X_n$ for all n. On the other hand, since the values of $X_n(x_{n,u})$ is decreasing, for each n there exists a $d(n)$ such that the first $d(n)$ terms of $X_n(x_{(n,i)})$ is $\geqslant \epsilon/2$ and the remaining $c(n) - d(n)$ terms are $< \epsilon/2$. As discussed informally before, for each n we consider the partition of space by $\bigvee_{i=1}^{d(n)} x_{n,i}$ and $\bigvee_{i>d(n)} x_{n,i}$. By the convergence assumption $X_n \to 0$, we must have

$$\bigwedge_{k=1}^{\infty} \bigvee_{i=1}^{d(k)} x_{k,i} = \varnothing$$

for no part of X_n can remain above ϵ forever. But we previously showed $\mathbb{P}$'s countable additivity is equivalent to continuity at $\mathbf{0}$, so $\mathbb{P}(\bigvee_{i=1}^{d(k)} x_{k,i}) \to 0$ as $k \to \infty$, which means for sufficiently large k, $\mathbb{P}(\bigvee_{i=1}^{d(k)} x_{k,i}) < \epsilon/2M$. Therefore, for sufficiently large n,

$$\begin{aligned}
\mathbb{E}X_n &= \sum_{i\geqslant 1} X_n(x_{n,i})\mathbb{P}(x_{n,i}) = \sum_{i=1}^{d(n)} \underbrace{X_n(x_{n,i})}_{\leqslant M} \mathbb{P}(x_{n,i}) + \sum_{i>d(n)} \underbrace{X_n(x_{n,i})}_{\leqslant \epsilon/2} \mathbb{P}(x_{n,i}) \\
&\leqslant M \sum_{i=1}^{d(n)} \mathbb{P}(x_{n,i}) + \frac{\epsilon}{2} \sum_{i>d(n)} \mathbb{P}(x_{n,i}) \\
&\leqslant M \cdot \mathbb{P}(\bigvee_{i=1}^{d(n)} x_{n,i}) + \frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.
\end{aligned}$$

END OF STEP 1

STEP 2: ELEMENTARY RANDOM VARIABLES. Now let $\{X_n\} \subset \mathfrak{E}$ be a sequence of elementary random variables, each possessing an expected value and converging downward to 0. The idea is to use STEP 1 to approximate these X_n's and show that the error term can be carefully controlled.

More formally, since $\mathbb{E}X_n = \sum_{i\geqslant 1} X_n(x_{n,i})\mathbb{P}(x_{n,i}) < \infty$ there exists a $e(n)$ such that $\sum_{i>e(n)} X_n(x_{n,i})\mathbb{P}(x_{n,i}) < \epsilon 2^{-n}$. The first $e(n)$ terms form a simple random variable $X_n' := \sum_{i=1}^{e(n)} X_n(x_{n,i}) I_{x_{(n.i)}}$. Further define $X_n'' := \bigwedge_{i=1}^{n} X_n'$ so that X_n'' is monotonically decreasing. Since $X_n \to 0$ and $0 \leqslant X_n'' \leqslant X_n' \leqslant X_n$, we know $X_n'' \to 0$ monotonically as well, and by the previous part $\mathbb{E}X_n'' \downarrow 0$.

How about the remainder? To fix the potential issue of X_n' not being monotone we introduced X_n'', but in doing so, controlling $X_n - X_n''$ becomes slightly more involved as well. We inductively prove that $X_n - X_n''$ can still be controlled. Of course, $X_1' = X_1 - X_1''$. For the inductive step, we use inclusion-exclusion on X_n'' and X_{n+1}:

$$\begin{aligned}
\mathbb{E}(X_{n+1}') + \mathbb{E}(X_n'') &= \mathbb{E}(X_{n+1}' \wedge X_n'') + \mathbb{E}(X_{n+1}' \vee X_n'') = \mathbb{E}(X_{n+1}'') + \mathbb{E}(X_{n+1}' \vee X_n'') \\
&\leqslant \mathbb{E}(X_{n+1}'') + \mathbb{E}(X_{n+1}' \vee X_n') \leqslant \mathbb{E}(X_{n+1}'') + \mathbb{E}X_n.
\end{aligned}$$

Rearranging gives

$$\mathbb{E}(X_{n+1}') \leqslant \mathbb{E}(X_{n+1}'') + \mathbb{E}X_n - \mathbb{E}X_n' < \mathbb{E}(X_{n+1}'') + \epsilon.$$

This shows $X_n - X_n''$ is not far from $X_n - X_n''$, so $\limsup \mathbb{E}X_n = \limsup(\mathbb{E}X_n'' + \mathbb{E}(X_n - X_n'')) \leqslant \epsilon$. Since ϵ is arbitrary the proof is complete. END OF STEP 2

STEP 3: (GENERAL) RANDOM VARIABLES. Let $\{X_n\} \subset \mathfrak{X}$ be L^1. For each n, let $\{Y_{n,k}\}_{k\geqslant 1}$ be a sequence of elementary random variables with expected values defined that converge uniformly to X_n. Further WLOG assume each $\{Y_{n,k}\}_{k\geqslant 1}$ is decreasing, for we may otherwise consider $\{\bigwedge_{i\geqslant k} Y_{n,k}\}_{k\geqslant 1}$. Consider the vertical sequence $Z_n = \bigwedge_{i=1}^n Y_{i,n}$. On one hand, since $Y_{n,\cdot} \downarrow X_n$ we must have $Z_n \geqslant X_n$. On the other hand, by construction we also know that $Y_{n,k} \geqslant Z_k$ if $n \leqslant k$. Finally, it is clear that Z_n is decreasing since by assumption $Y_{n,k} \geqslant Y_{n,k+1}$. Combining these three identities along with $X_n \to 0$, we see that $\lim Z_n = \lim X_n = 0$. But Z_n is elementary, so by PART 2 $\mathbb{E}Z_n \downarrow 0$. Finally, noting that $Z_n \geqslant X_n$ so $\mathbb{E}Z_n \geqslant \mathbb{E}X_n$, we conclude that $\mathbb{E}X_n \downarrow 0$. □

A direct consequence of DCT is that if $X_n \to X$ is monotonically decreasing and L^1, then $\mathbb{E}X_n \downarrow \mathbb{E}X$ since we can apply the previous proof to $X_n - X$, a sequence of L^1 random variables converging downward to 0. This also gives rise to the:

Theorem: Monotone Convergence Theorem

If $\{X_n\} \subset L^1$ is monotone increasing and $X_n \to X \in L^1$, then $\mathbb{E}X_n \uparrow \mathbb{E}X$.

Finally, we prove the

Theorem: Fatou's Lemma

Let $\{X_n\} \subset L^1$. If there exists a uniform bound $M \in L^1$ such that $0 \leqslant X_n \leqslant M$ then

$$\mathbb{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathbb{E}X_n,$$

where $\liminf_{n\to\infty} X_n$ is defined as $\bigvee_{k=1}^{\infty} \bigwedge_{n\geqslant k} X_n$.

Proof. First, $\liminf X_n$ is well-defined because $\bigwedge_{n\geqslant k} X_n$ is bounded for each k, and $\mathfrak{X}$ is closed under arbitrary joins and meets taken over bounded subsets. Further, since $0 \leqslant \liminf X_n \leqslant M$ we see it is also in L^1. On the other hand note that $\bigwedge_{n\geqslant k} X_n$ is strictly increasing to $\liminf_n X_n$ as $n \to \infty$ and is uniformly bounded by M, so by MCT, $\mathbb{E}(\bigwedge_{n\geqslant k} X_n) \uparrow \mathbb{E}(\liminf_n X_n)$. But by definition $\bigwedge_{n\geqslant k} X_n \leqslant X_k$ for each k, so $\mathbb{E}(\bigwedge_{n\geqslant k} X_n) \leqslant \mathbb{E}X_k$ for each k, and we conclude that

$$\mathbb{E}(\liminf_{n\to\infty} X_n) = \lim_{k\to\infty} \mathbb{E}(\bigwedge_{n\geqslant k} X_n) \leqslant \liminf_{n\to\infty} \mathbb{E}X_n.$$ □