Probability Theory

February 10, 2019

1 Notation

$\mathbb{R}$	: Real line.
$\mathbb{R}^*$	: Extended real line, *i.e.*, $\mathbb{R}^* := \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.
$\mathbb{R}^*_+$	: Non-negative extended real line, *i.e.*, $\mathbb{R}^*_+ := \{r \in \mathbb{R}^*; r \geq 0\}$.
$(a_n) \uparrow a$, for $a_n, a \in \mathbb{R}^*$	: (a_n) is a monotonically increasing real (extended) sequence (*i.e.*, $a_{n+1} \geq a_n$, $\forall n$) and (a_n) converges to a.
$(f_n) \uparrow f$, for $f, f_n : \Omega \to \mathbb{R}^*$	: (f_n) is a monotonically increasing real (extended) valued function sequence (*i.e.*, $f_{n+1}(\omega) \geq f_n(\omega)$, $\omega \in \Omega$) and (f_n) converges to f, *i.e.*, $\lim_{n\to\infty} f_n(\omega) = f(\omega), \forall \omega \in \Omega$.
I_A	: Indicator function, *i.e.*, $I_A = 1$ if $\omega \in A$ and $I_A = 0$ otherwise.
$f_1 \wedge f_2$, for $f_1, f_2 : \Omega \to \mathbb{R}^*$	: $f \wedge f_2$ is a function from Ω to $\mathbb{R}^*$ defined as $(f_1 \wedge f_2)(\omega) = \min\{f_1(\omega), f_2(\omega)\}$.
$f_1 \vee f_2$, for $f_1, f_2 : \Omega \to \mathbb{R}^*$	: $f \vee f_2$ is a function from Ω to $\mathbb{R}^*$ defined as $(f_1 \vee f_2)(\omega) = \max\{f_1(\omega), f_2(\omega)\}$.

2 Probability space

Definition: The 3-tuple $(\Omega, \mathcal{F}, P)$ is called a probability space, where

1. Ω is a set called the sample space.

2. $\mathcal{F}$ is a σ-field.

   Definition of σ-field: $\mathcal{F}$ is a non-empty collection of subsets of Ω which satisfies

   (S1) $\Omega \in \mathcal{F}$.

   (S2) If $A \subseteq \Omega$ and $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.

   (S3) If each set in the collection $\{A_n; n \in \mathbb{N}\}$ belongs to $\mathcal{F}$, *i.e.*, $A_n \in \mathcal{F}$, $\forall n \in \mathbb{N}$ (not necessarily disjoint), then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

   Note that $A \subseteq \Omega$ is called $\mathcal{F}$-set if $A \in \mathcal{F}$.

3. P is a probability measure.

Definition of probability measure: $P : \mathcal{F} \to [0,1]$ is called a probability measure if it satisfies:

(M1) $P(\Omega) = 1$ and $P(\emptyset) = 0$.

(M2) If $\{A_n\}_{n\in\mathbb{N}}$ is a disjoint collection of $\mathcal{F}$-sets, *i.e.*, $A_k \cap A_j = \emptyset$, for $k \neq j$, then

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n). \tag{1}$$

This property is called the *countable additivity of the probability measure.*

In other words, P is a set function (*i.e.*, P takes sets in $\mathcal{F}$ to real values in $[0,1]$) which satisfies M1 and M2.

Remark 1. *A similar concept to countable additivity is the finite additivity which is defined as follows: If $\{A_i; 1 \leq i \leq n\}$ is a finite collection of disjoint $\mathcal{F}$-sets, then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$. Note that countable additivity implies finite additivity. Indeed, by considering the countable collection $\{B_i; i \in \mathbb{N}\}$, where $B_1 = A_1, \ldots, B_n = A_n$, and $B_k = \emptyset$, for $k > n$, the claim follows.*

Remark 2. *A more generalized set function is the notion of measure. A measure $\mu : \mathcal{F} \to \mathbb{R}_+^*$ (contrary to the probability measure where the range of P is contained in $[0,1]$) which satisfies $\mu(\emptyset) = 0$ (need not satisfy $\mu(\Omega) = 1$) and countable additivity (M2). Thus, probability measure is a measure with the additional condition that $P(\Omega) = 1$.*

Lemma 1. *If A and B are $\mathcal{F}$-sets with $A \subseteq B$, then $P(A) \leq P(B)$. Also, $P(B \setminus A) = P(B) - P(A)$.*

Proof. Note that since $A \subseteq B$, we have $B = A \cup (B \setminus A)$ and, A and $B \setminus A$ are disjoint. Now, by the finite additivity of P, we have

$$P(B) = P(A) + \underbrace{P(B \setminus A)}_{\geq 0} \tag{2}$$

$$\Rightarrow P(B) \geq P(A).$$

This proves the first part. The second part follows from Eq. (2). ☐

Lemma 2. *If A is an $\mathcal{F}$-set, then $P(A^c) = 1 - P(A)$.*

Proof. Note that $A \cup A^c = \Omega$. Also, A and A^c are disjoint. Therefore by finite additivity property of P and M1, we have

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c).$$

Hence, the claim follows. ☐

2.1 Limit of sets

Definition: (Liminf of a sequence of sets) Given a sequence of sets $(A_n)_{n\in\mathbb{N}}$, where $A_n \subseteq \Omega$, we define

$$\liminf_n A_n = \bigcup_n \bigcap_{k\geq n} A_k. \tag{3}$$

Definition: (Limsup of a sequence of sets) Given a sequence of sets $(A_n)_{n\in\mathbb{N}}$, where $A_n \subseteq \Omega$, we define

$$\limsup_n A_n = \bigcap_n \bigcup_{k\geq n} A_k. \tag{4}$$

Definition: (Limit of a sequence of sets) We say the limit of the sequence of sets $(A_n)_{n\in\mathbb{N}}$ exists if $\liminf_n A_n = \limsup_n A_n$ and the $\lim_n A_n$ is that common set.

We will consider specific sequences here

2.1.1 Monotonically increasing sequence of sets

Definition: A sequence $(A_n)_{n\in\mathbb{N}}$ is called *monotonically increasing* sequence if $A_n \subseteq A_{n+1}$, $\forall n \in \mathbb{N}$.

In this case, note that for $n \in \mathbb{N}$,

$$\bigcap_{k\geq n} A_k = A_n, \text{ since } A_n \subseteq A_{n+1} \subseteq A_{n+2} \ldots$$

Therefore,

$$\liminf_n A_n = \bigcup_n \bigcap_{k\geq n} A_k = \bigcup_n A_n. \tag{5}$$

Note that for $n > 1$, since $A_1 \subseteq A_2 \cdots \subseteq A_{n-1} \subseteq A_n$, we have

$$\bigcup_{k=1}^{n} A_k = A_n \Rightarrow \bigcup_{k\geq n} A_k = \bigcup_{k\geq 1} A_k \tag{6}$$

Therefore,

$$\limsup_n A_n = \bigcap_n \bigcup_{k\geq n} A_k = \bigcap_n \bigcup_{k\geq 1} A_k = \bigcup_{k\geq 1} A_k. \tag{7}$$

Therefore, by the definition of $\lim_n A_n$, we have

$$\lim_n A_n = \bigcup_n A_n. \tag{8}$$

The next question is what happens to the probability of the monotonically increasing sets A_n when each A_n is an $\mathcal{F}$-set. Indeed, we are considering the real sequence $(P(A_n))_{n\in\mathbb{N}}$. The real sequence $(P(A_n))_{n\in\mathbb{N}}$ is bounded since $0 \le P(A_n) \le 1$, $\forall n \in \mathbb{N}$. Also since the set sequence $(A_n)_{n\in\mathbb{N}}$ is monotonically increasing, we have, for $n \in \mathbb{N}$,

$$A_{n+1} \supseteq A_n \Rightarrow P(A_{n+1}) \ge P(A_n), \text{ (follows from Lemma 1)}$$

Therefore, the real sequence $(P(A_n))_{n\in\mathbb{N}}$ is a monotonically increasing bounded sequence. Hence it should converge. But where does it converges to?

Theorem 1. *If $(A_n)_{n\in\mathbb{N}}$ is a monotonically increasing sequence of $\mathcal{F}-$sets, then*

$$\lim_{n\to\infty} P(A_n) = P(\lim_n A_n) = P(\bigcup_n A_n). \tag{9}$$

Proof. Let $A_0 = \emptyset$. Now set

$$\begin{aligned} B_1 &:= A_1 \setminus A_0; \\ B_2 &:= A_1 \setminus A_1; \\ &\vdots \\ B_n &:= A_n \setminus A_{n-1}; \\ &\vdots \end{aligned}$$

Now note that the set sequence $(B_n)_{n\in\mathbb{N}}$ is a disjoint sequence, *i.e.*, $B_i \cap B_j = \emptyset$, for $i \ne j$. Also,

$$\bigcup_n B_n = \bigcup_n A_n. \tag{10}$$

Therefore, from Eq. (10) and the fact that the set sequence $(A_n)_{n\in\mathbb{N}}$ is monotonically increasing, we have

$$\begin{aligned} P(\lim_n A_n) = P(\bigcup_n A_n) &= P(\bigcup_n B_n) \\ &= \sum_{n\in\mathbb{N}} P(B_n) \text{ (follows from M2)} \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i) - P(A_{i-1}) \text{ (follows from Lemma 1)} \\ &= \lim_{n\to\infty} P(A_n) - \underbrace{P(A_0)}_{=0} \\ &= \lim_{n\to\infty} P(A_n). \end{aligned}$$

□

Remark 3. *Note that in the proof of the above theorem, we never used the condition $P(\Omega) = 1$ of the probability measure. This implies that the above result also holds for any measure on Ω.*

2.1.2 Monotonically decreasing sequence of sets

Definition: A sequence $(A_n)_{n\in\mathbb{N}}$ is called *monotonically decreasing* sequence if $A_{n+1} \subseteq A_n$, $\forall n \in \mathbb{N}$.

In this case, note that for $n \in \mathbb{N}$,

$$\cup_{k\geq n} A_k = A_n, \text{ since } A_n \supseteq A_{n+1} \supseteq A_{n+2} \ldots$$

Therefore,

$$\limsup_n A_n = \bigcap_n \bigcup_{k\geq n} A_k = \bigcap_n A_n. \tag{11}$$

Note that for $n > 1$, since $A_1 \supseteq A_2 \cdots \supseteq A_{n-1} \supseteq A_n$, we have

$$\bigcap_{k=1}^{n} A_k = A_n \Rightarrow \bigcap_{k\geq n} A_k = \bigcap_{k\geq 1} A_k \tag{12}$$

Therefore,

$$\liminf_n A_n = \bigcup_n \bigcap_{k\geq n} A_k = \bigcup_n \bigcap_{k\geq 1} A_k = \bigcap_{k\geq 1} A_k. \tag{13}$$

Therefore, by the definition of $\lim_n A_n$, we have

$$\lim_n A_n = \bigcap_n A_n. \tag{14}$$

What happens to the probability of the monotonically decreasing sets A_n when each A_n is an $\mathcal{F}$-set. Here also, the real sequence $(P(A_n))_{n\in\mathbb{N}}$ is bounded since $0 \leq P(A_n) \leq 1$, $\forall n \in \mathbb{N}$. Also since the set sequence $(A_n)_{n\in\mathbb{N}}$ is monotonically decreasing, we have, for $n \in \mathbb{N}$,

$$A_{n+1} \subseteq A_n \Rightarrow P(A_{n+1}) \leq P(A_n), \text{ (follows from Lemma 1)}$$

Therefore, the real sequence $(P(A_n))_{n\in\mathbb{N}}$ is a monotonically decreasing bounded sequence. Hence it should converge.

Theorem 2. *If $(A_n)_{n\in\mathbb{N}}$ is a monotonically decreasing sequence of $\mathcal{F}-$sets, then*

$$\lim_{n\to\infty} P(A_n) = P(\lim_n A_n) = P(\bigcap_n A_n). \tag{15}$$

Proof. Since $(A_n)_{n\in\mathbb{N}}$ is a monotonically decreasing sequence of $\mathcal{F}-$sets, we have $(A_n^c)_{n\in\mathbb{N}}$ to be a monotonically increasing sequence of $\mathcal{F}-$sets. This follows from S2.

Now from Theorem 1, we know that

$$\lim_{n\to\infty} P(A_n^c) = P(\lim_n A_n^c) = P(\bigcup_n A_n^c) \tag{16}$$

However, note that $\bigcup_n A_n^c = (\cap_n A_n)^c$. Therefore from Lemma 2 and Eq. (16), we have

$$\begin{aligned}
&\lim_{n\to\infty} 1 - P(A_n) = 1 - P(\bigcap_n A_n) \\
&\Leftrightarrow 1 - \lim_{n\to\infty} P(A_n) = 1 - P(\bigcap_n A_n) \\
&\Leftrightarrow \lim_{n\to\infty} P(A_n) = P(\bigcap_n A_n).
\end{aligned}$$

□

3 Random variables

Definition: (Borel σ-field) The smallest σ-field on $\mathbb{R}^*$ containing intervals. Recall that intervals are of the form $(a, b), [a, b], [a, b), (a, b]$, where $a, b \in \mathbb{R}^*$ and $a \leq b$.

Remark 4. *The definition is indeed well-defined. Note that given a collection C of subsets of $\mathbb{R}^*$, one can ask what is the smallest σ-field containing C. We denote such a sigma field as $\sigma(C)$. Indeed, one can obtain $\sigma(C)$ as follows. Consider the new collection $\mathcal{G} := \{\mathcal{H}$ s.t. $\mathcal{H}$ is a σ-field and $C \subseteq \mathcal{H}\}$. Note that this is a collection of σ-fields. Is $\mathcal{G}$ non-empty? YES - since the power set of $\mathbb{R}^*$ itself is a σ-field and it contains C. Hence the power set belongs to $\mathcal{G}$. Now it is easy to verify that*

$$\sigma(C) = \bigcap_{\mathcal{H}\in\mathcal{G}} \mathcal{H}. \tag{17}$$

Definition: (Random variable) A function $X : \Omega \to \mathbb{R}^*$ is called a *random variable (r.v.)* if $X^{-1}(B) \in \mathcal{F}$, for every $B \in \mathcal{B}$. Here, $X^{-1}(B)$ is defined as follows: for $B \subseteq \mathbb{R}^*$,

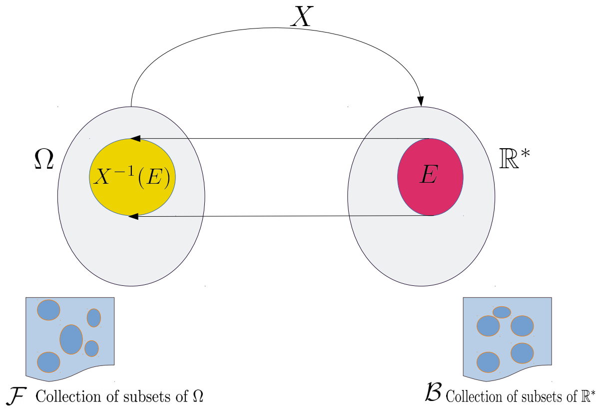
$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}. \tag{18}$$

By the above it is hard to verify whether a function $X : \omega \to \mathbb{R}^*$ is a r.v. since we don't know the sets inside $\mathcal{B}$. However, we do know that the intervals are inside $\mathcal{B}$. However, the following claim reduces this effort by providing a sufficient condition.

Theorem 3. *If $X^{-1}([-\infty, a]) \in \mathcal{F}$, $\forall a \in \mathbb{R}^*$, then X is a r.v.*

Proof. Given that $X^{-1}([-\infty, a]) \in \mathcal{F}$, $\forall a \in \mathbb{R}^*$, we have to show that X is a r.v. Define

$$\mathcal{C} := \{B \subseteq \mathbb{R}^* | X^{-1}(B) \in \mathcal{F}\} \tag{19}$$



If we can show that $\mathcal{B} \subseteq \mathcal{C}$ we are done. Because if so then for every $E \in \mathcal{B}$, we have $X^{-1}(E) \in \mathcal{F}$ (by definition of $\mathcal{C}$). To do so we show that $\mathcal{C}$ is a $\sigma-$field containing intervals. Since $\mathcal{B}$ (the Borel σ-field) is the smallest σ-field containing intervals, we have $\mathcal{B} \subseteq \mathcal{C}$.

Part 1: To show that $\mathcal{C}$ contains intervals
From the hypothesis we know that $[-\infty, a] \in \mathcal{C}$, $\forall a \in \mathbb{R}$. Now note that for $b \in \mathbb{R}^*$, we have

$$[-\infty, b) = \bigcup_{n \in \mathbb{N}} [-\infty, b - \frac{1}{n}]. \tag{20}$$

Therefore,

$$\begin{aligned} X^{-1}([-\infty, b)) &= X^{-1}\left(\bigcup_{n \in \mathbb{N}} [-\infty, b - \frac{1}{n}]\right) \\ &= \underbrace{\bigcup_{n \in \mathbb{N}} \underbrace{X^{-1}([-\infty, b - \frac{1}{n}])}_{\in \mathcal{F} \text{ by hypothesis}}}_{\in \mathcal{F} \text{ by countable union}} \end{aligned}$$

$$\bullet \text{ This implies that } [-\infty, b) \in \mathcal{C},\ \forall b \in \mathbb{R}^*. \tag{21}$$

Now note that

$$X^{-1}((b, +\infty]) = X^{-1}\left([-\infty, b]^c\right) = \underbrace{\left(X^{-1}([-\infty, b])\right)^c}_{\in \mathcal{F} \text{ since } X^{-1}([-\infty,b]) \in \mathcal{F} \text{ by hypothesis}}$$

$$\bullet \text{ This implies that } (b, +\infty] \in \mathcal{C},\ \forall b \in \mathbb{R}^*. \tag{22}$$

Also, note that

$$X^{-1}([b,+\infty]) = X^{-1}\left([-\infty,b)^c\right) = \underbrace{\left(X^{-1}([-\infty,b)\right)^c}_{\in\mathcal{F} \text{ since } X^{-1}([-\infty,b))\in\mathcal{F} \text{ by Eq. (21)}}$$

- This implies that $(b,+\infty] \in \mathcal{C}$, $\forall b \in \mathbb{R}^*$. (23)

Further, for $a,b \in \mathbb{R}^*$, $a < b$, we have

$$(a,b) = (a,+\infty] \cap [-\infty,b) \Rightarrow X^{-1}((a,b)) = \underbrace{X^{-1}((a,+\infty])}_{\in\mathcal{F} \text{ Eq. (22)}} \cap \underbrace{X^{-1}([-\infty,b))}_{\in\mathcal{F} \text{ Eq. (21)}}.$$

- This implies that $(a,b) \in \mathcal{C}$, $\forall a, \forall b \in \mathbb{R}^*, a < b$. (24)
- Similarly, $[a,b), [a,b], (a,b] \in \mathcal{C}$, $\forall a, \forall b \in \mathbb{R}^*, a < b$. (25)

Part 2: To show that $\mathcal{C}$ is a σ-field over $\mathbb{R}^*$
Note that $X^{-1}(\mathbb{R}^*) = \Omega \in \mathcal{F}$. Therefore,

$$\mathbb{R}^* \in \mathcal{C}. \tag{26}$$

If $A \in \mathcal{C}$, then $X^{-1}(A) \in \mathcal{F}$. Therefore,

$$\begin{aligned} &X^{-1}(A^c) = \left(X^{-1}(A)\right)^c \in \mathcal{F} \\ &\Rightarrow A^c \in \mathcal{C}. \end{aligned} \tag{27}$$

Given a countable collection $\{A_n\}_{n\in\mathbb{N}}$ with $A_n \in \mathcal{C}$, $\forall n \in \mathbb{N}$ (which implies that $X^{-1}(A_n) \in \mathcal{F}, \forall n$ by the definition of $\mathcal{C}$), we have

$$\begin{aligned} X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \underbrace{\bigcup_{n=1}^{\infty} \underbrace{X^{-1}(A_n)}_{\in\mathcal{F}}}_{\in\mathcal{F}} \\ \Rightarrow \bigcup_{n=1}^{\infty} A_n &\in \mathcal{C}. \end{aligned} \tag{28}$$

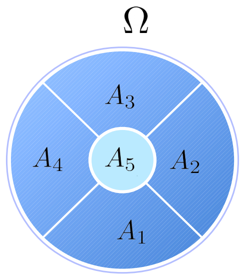
Therefore, $\mathcal{C}$ is a σ-field over $\mathbb{R}^*$. $\square$

We now consider the simplest of random variables.

3.1 Non-negative simple functions

Definition: We call a finite collection $\{A_i\}_{i=1}^n$ an $\mathcal{F}$-partition of Ω if

1. Each $A_i \in \mathcal{F}$.
2. A_i's are disjoint (*i.e.*, $A_k \cap A_t = \emptyset$, if $k \neq t$) and
3. $\bigcup_{i=1}^n A_i = \Omega$ (*i.e.* their union gives the entire set Ω).



Definition: A function $s : \Omega \to \mathbb{R}_+^*$ is called a non-negative simple function if it has the form

$$s(\omega) = \sum_{i=1}^{n} a_i I_{A_i}(\omega), \text{ where } a_i \in \mathbb{R}_+^*, 1 \le \forall i \le n. \tag{29}$$

Note that s is a $r.v.$ To see that, lets assume that $a_1 < a_2 < a_3 < \cdots < a_n$ (if not, then re-number). Then

$$s^{-1}([-\infty, a]) = \begin{cases} \emptyset, \text{ if } a < a_1. \\ A_1, \text{ if } a_1 \le a < a_2. \\ A_1 \cup A_2, \text{ if } a_2 \le a < a_3. \\ A_1 \cup A_2 \cup A_3, \text{ if } a_3 \le a < a_4. \\ \vdots \\ \Omega, \text{ if } a \ge a_n. \end{cases}$$

Thus $s^{-1}([-\infty, a]) \in \mathcal{F}$, $\forall a \in \mathbb{R}^*$. Therefore s is a $r.v.$

We denote by $\mathbb{L}_0^+$ the collection of non-negative simple functions.

$$\mathbb{L}_0^+ := \{s : \Omega \to \mathbb{R}_+^* | s \text{ is a non-negative simple function}\}. \tag{30}$$

Properties:

Proposition 1. *If $s_1, s_2 \in \mathbb{L}_0^+$, then*

1. $s_1 + s_2 \in \mathbb{L}_0^+$ *and* $s_1 s_2 \in \mathbb{L}_0^+$.
2. $cs_1 \in \mathbb{L}_0^+$, *for* $c \in \mathbb{R}_+^*$.
3. $s_1 \wedge s_2 \in \mathbb{L}_0^+$.
4. $s_1 \vee s_2 \in \mathbb{L}_0^+$.

Proof. Let

$$s_1 = \sum_{i=1}^{n} a_i I_{A_i} \text{ and } s_2 = \sum_{j=1}^{m} b_j I_{B_j}.$$

1. It is easy to verify that $\{A_i \cap B_j; 1 \leq i \leq n, 1 \leq j \leq m\}$ is a $\mathcal{F}$-partition. Then

$$s_1 + s_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j)\, I_{A_i \cap B_j}. \tag{31}$$

   To justify this claim, note that

$$\begin{aligned}
&\text{For } \omega \in \Omega \Rightarrow \omega \in A_i \text{ and } \omega \in B_j,\ \text{for some } i, j, 1 \leq i \leq n, 1 \leq j \leq m, \\
&\qquad \text{since } \{A_i\}, \{B_j\} \text{ are } \mathcal{F} - \text{partitions.} \\
&\Leftrightarrow \omega \in A_i \cap B_j \\
&\Leftrightarrow s_1(\omega) = a_i \text{ and } s_2(\omega) = b_j \text{ with } \omega \in A_i \cap B_j \\
&\Leftrightarrow (s_1 + s_2)(\omega) = s_1(\omega) + s_2(\omega) = a_i + b_j,\ \text{with } \omega \in A_i \cap B_j \\
&\Leftrightarrow s_1 + s_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j)\, I_{A_i \cap B_j}.
\end{aligned}$$

   Therefore $s_1 + s_2 \in \mathbb{L}_0^+$.

2. Similarly, $s_1 s_2 \in \mathbb{L}_0^+$ with

$$s_1 s_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j I_{A_i \cap B_j}. \tag{32}$$

3. Also, for $c \in \mathbb{R}_+^*$, $cs_1 \in \mathbb{L}_0^+$ with

$$cs_1 = \sum_{i=1}^{n} \sum_{j=1}^{m} c a_i I_{A_i}. \tag{33}$$

4. $s_1 \wedge s_2 \in \mathbb{L}_0^+$ with

$$s_1 \wedge s_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \min\{a_i, b_j\} I_{A_i \cap B_j}. \tag{34}$$

5. $s_1 \vee s_2 \in \mathbb{L}_0^+$ with

$$s_1 \vee s_2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \max\{a_i, b_j\} I_{A_i \cap B_j}. \tag{35}$$

□

The simple functions even though are simple are not that simple. They are strong enough to approximate any non-negative *r.v.*

Theorem 4. *If X is a non-negative r.v., then there exists a sequence (s_n), where $s_n \in \mathbb{L}_0^+$ s.t. $s_n \uparrow X$. This means that for each $\omega \in \Omega$, we have $(s_n(\omega))$ is a monotonically increasing sequence and $\lim_{n\to\infty} s_n(\omega) = X(\omega)$.*

Proof. We will create the sequence (s_n) as follows: Let

$$E_{n,k} := \Big[\frac{k-1}{2^n}, \frac{k}{2^n}\Big), 1 \leq k \leq n2^n \text{ and } E_{n,\infty} = [n, +\infty]. \tag{36}$$

Also, let

$$A_{n,k} := X^{-1}\left(E_{n,k}\right), 1 \leq k \leq n2^n \text{ and } A_{n,\infty} = X^{-1}(E_{n,\infty}). \tag{37}$$

Define

$$s_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} I_{A_{n,k}} + nI_{A_{n,\infty}}. \tag{38}$$

It is easy to verify that $s_n \in \mathbb{L}_0^+$ since $\{A_{n,k}, 1 \leq k \leq n2^n; A_{n,\infty}\}$ is an $\mathcal{F}$-partition.

It is also easy to verify from Fig. 2 that

$$s_{n+1}(\omega) \geq s_n(\omega), \forall \omega \in \Omega. \tag{39}$$

Now we will verify that $\lim_{n\to\infty} s_n(\omega) = X(\omega)$, $\forall \omega \in \Omega$.

For $\omega \in \Omega$, there are two cases possible
1) Either $\omega \in A_{n,k}$ for some $1 \leq k \leq n2^n$. In this case,

$$\begin{aligned}
&s_n(\omega) = \frac{k-1}{2^n} \text{ and } X(\omega) \in E_{n,k} \\
&\Rightarrow \frac{k-1}{2^n} \leq X(\omega) < \frac{k}{2^n} \\
&\Rightarrow \frac{k-1}{2^n} - \frac{k-1}{2^n} \leq X(\omega) - s_n(\omega) < \frac{k}{2^n} - \frac{k-1}{2^n} \\
&\Rightarrow 0 \leq X(\omega) - s_n(\omega) < \frac{1}{2^n}. \\
&\Rightarrow \lim_{n\to\infty} s_n(\omega) = X(\omega) \text{ (by squeeze theorem)}.
\end{aligned}$$

2) Or $\omega \in A_{n,\infty}$. In this case, we have

$$\begin{aligned}
&s_n(\omega) = n \text{ and } X(\omega) \in [n, +\infty] \\
&\Rightarrow s_n(\omega) = n \text{ and } X(\omega) \geq n.
\end{aligned}$$

Hence, we cannot obtain the bound similar to the earlier case. However, one can consider two sub-cases here: 1) If $X(w) < +\infty$. In this case, by the Archimedean theorem, there exists an $N \in \mathbb{N}$ s.t. $N > X(\omega)$. Therefore, $\forall n \geq N$, we have the bound

$$\Rightarrow 0 \leq X(\omega) - s_n(\omega) < \frac{1}{2^n}. \tag{40}$$

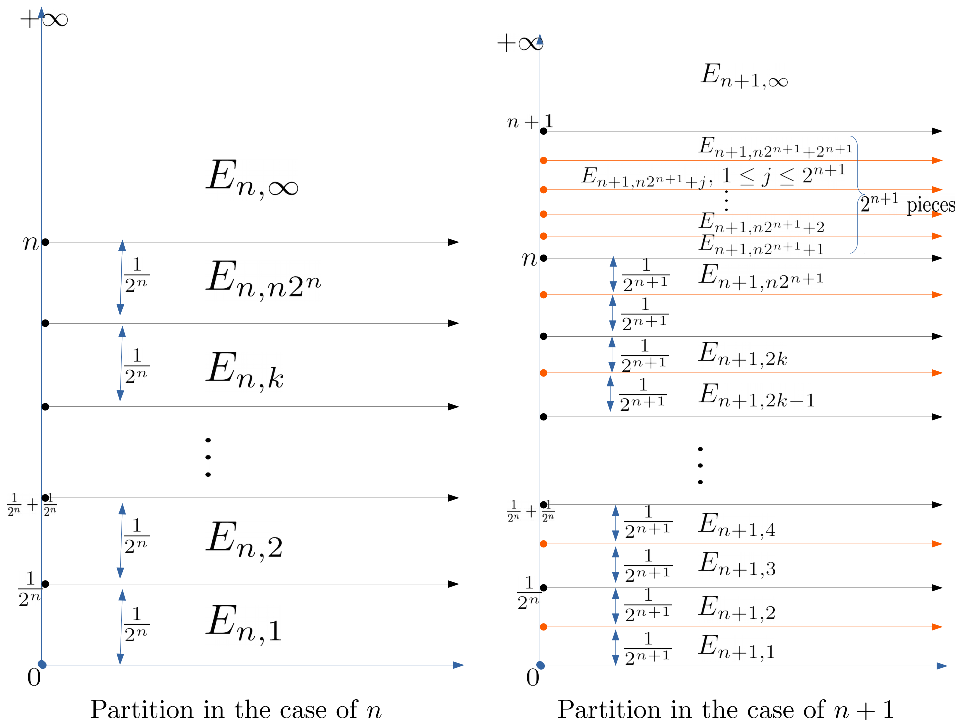


Figure 1: Partitions

Therefore, $\lim_{n\to\infty} s_n(\omega) = X(\omega)$, by squeeze theorem.
2) If If $X(w) = +\infty$. In this case, we have $s_n(\omega) = n$. Therefore,

$$\lim_{n\to\infty} s_n(\omega) = +\infty = X(\omega).$$

Thus, we have addressed every possible scenario. Therefore,

$$\lim_{n\to\infty} s_n(\omega) = X(\omega), \forall \omega \in \Omega. \tag{41}$$

$\square$

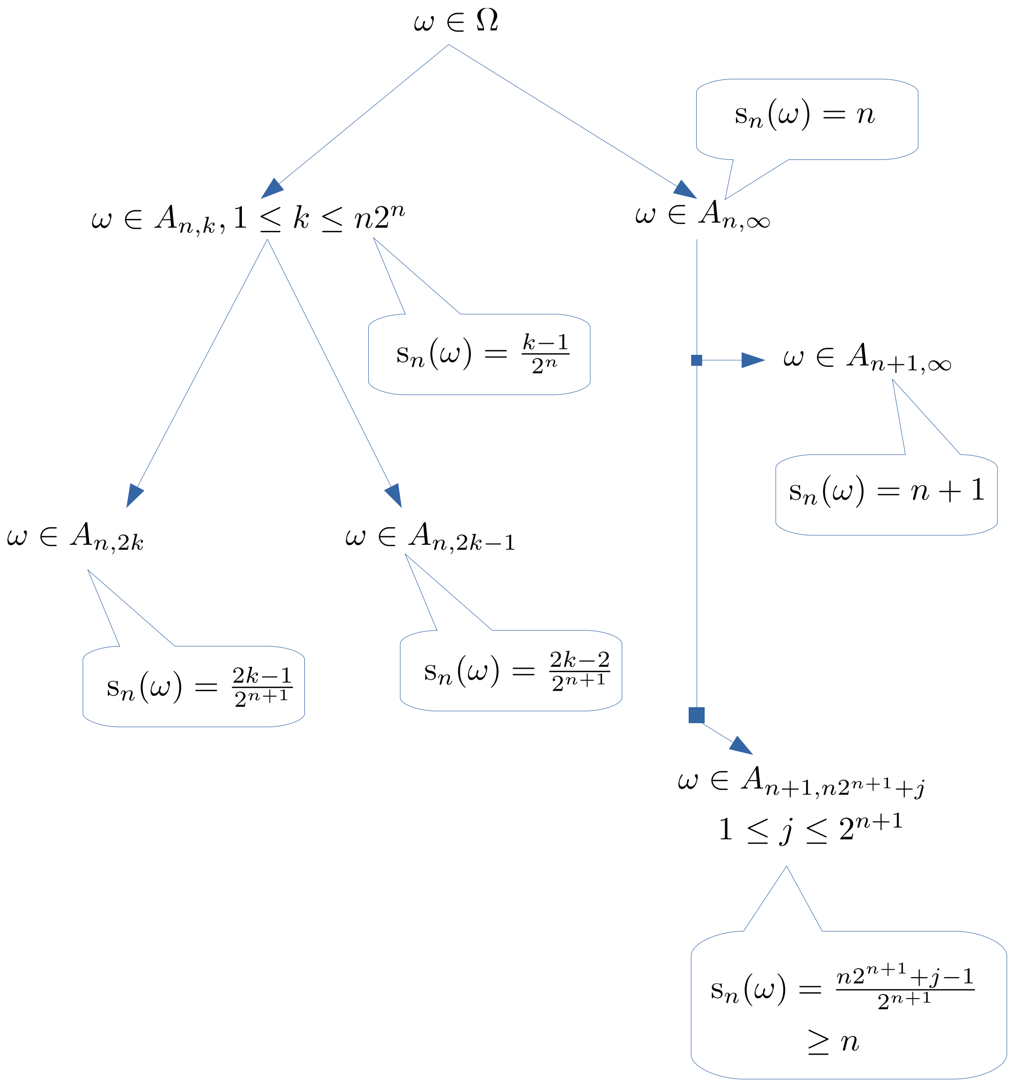
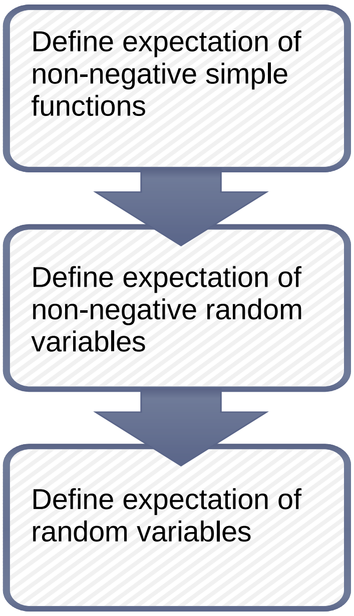


Figure 2: Illustration to show that $s_{n+1} \geq s_n$

4 Expectation of a random variable

Goal:



We first define the expectation of the non-negative simple functions as follows: For $s \in \mathbb{L}_0^+$ with $s = \sum_{i=1}^n a_i I_{A_i}$, ($\{A_i\}$ is an $\mathcal{F}$-partition and $a_i \in \mathbb{R}_+^*$), we define

$$\mathbb{E}\left[s\right] = \sum_{i=1}^{n} a_i P(A_i). \tag{42}$$

Properties of expectation of non-negative simple functions

Theorem 5. *For $s_1, s_2 \in \mathbb{L}_0^+$ with $s_1 = \sum_{i=1}^n a_i I_{A_i}$ and $s_2 = \sum_{j=1}^m b_j I_{B_j}$, ($\{A_i; 1 \leq i \leq n\}$ and $\{B_j; 1 \leq j \leq m\}$ are $\mathcal{F}$-partitions and $a_i, b_j \in \mathbb{R}_+^*$) , we have*

1. $\mathbb{E}[s_1] \geq 0$.
2. $\mathbb{E}[s_1 + s_2] = \mathbb{E}[s_1] + \mathbb{E}[s_2]$.
3. *For $c \in \mathbb{R}_+^*$, $\mathbb{E}[cs_1] = c\mathbb{E}[s_1]$.*
4. *If $s_1 \geq s_2$, then $\mathbb{E}[s_1] \geq \mathbb{E}[s_2]$. (Note that $s_1 \geq s_2$ means that $s_1(\omega) \geq s_2(\omega), \forall \omega \in \Omega$)*

Proof. 1.

$$\begin{aligned}\mathbb{E}\left[s_1\right] &= \sum_{i=1}^{n} \underbrace{a_i}_{\geq 0} \underbrace{P(A_i)}_{\geq 0} \\ &\geq 0.\end{aligned}$$

2. We know that

$$\begin{aligned} s_1 + s_2 &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j)\, I_{A_i \cap B_j} \\ \Rightarrow \mathbb{E}[s_1 + s_2] &= \mathbb{E}\left[\sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j)\, I_{A_i \cap B_j}\right] \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j)\, P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i P(A_i \cap B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} b_j P(A_i \cap B_j) \\ &= \sum_{i=1}^{n} a_i P(A_i \cap (\cup_{j=1}^{m} B_j)) + \sum_{j=1}^{m} b_j P((\cup_{i=1}^{n} A_i) \cap B_j) \text{ (by M2)} \\ &= \sum_{i=1}^{n} a_i P(A_i \cap \Omega) + \sum_{j=1}^{m} b_j P(\Omega \cap B_j) \\ &= \sum_{i=1}^{n} a_i P(A_i) + \sum_{j=1}^{m} b_j P(B_j) \\ &= \mathbb{E}\left[s_1\right] + \mathbb{E}\left[s_2\right]. \end{aligned}$$

3. Again,

$$\begin{aligned} cs_1 &= \sum_{i=1}^{n} c a_i I_{A_i} \\ \Rightarrow \mathbb{E}\left[cs_1\right] &= \sum_{i=1}^{n} c a_i P(A_i) = c \sum_{i=1}^{n} a_i P(A_i) = c\mathbb{E}\left[s_1\right]. \end{aligned}$$

4. For $s_1 \geq s_2$, we have

$$\begin{aligned}
\mathbb{E}[s_1] &= \mathbb{E}\left[\sum_{i=1}^{n} a_i I_{A_i}\right] \\
&= \sum_{i=1}^{n} a_i P(A_i \cap \Omega) \\
&= \sum_{i=1}^{n} a_i P(A_i \cap (\cup_{j=1}^{m} B_j)) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i P(A_i \cap B_j) \text{ (by M2)} \\
&= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ A_i \cap B_j \neq \emptyset}}^{m} a_i P(A_i \cap B_j) \\
&\geq \sum_{i=1}^{n} \sum_{\substack{j=1 \\ A_i \cap B_j \neq \emptyset}}^{m} b_j P(A_i \cap B_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} b_j P(A_i \cap B_j) \\
&= \mathbb{E}\left[s_2\right].
\end{aligned}$$

$\square$

Lemma 3. *Let* $(s_n) \uparrow s$*, where* $s_n, s \in \mathbb{L}_0^+$*. Then* $(\mathbb{E}[s_n]) \uparrow \mathbb{E}[s]$*.*