# HOL Theorem Proving and Formal Probability (4)

Chun TIAN
chun.tian@anu.edu.au

10/04/2024

# Countable Additivity of $(\mathbb{R}, \mathcal{S}, \lambda_0)$ (1)

Let $I_n = [a_n, b_n)$ and $\dot\bigcup_{n\in\mathbb{N}} I_n = [a, b)$, we prove:

$$(1)\ \sum_{n=1}^{\infty} \lambda_0(I_n) \leqslant \lambda_0(\dot\bigcup_{n\in\mathbb{N}} I_n) \quad \text{and} \quad (2)\ \lambda_0(\dot\bigcup_{n\in\mathbb{N}} I_n) \leqslant \sum_{n=1}^{\infty} \lambda_0(I_n)$$

Proof of subgoal 1 (easy):

$$\forall N.\ \sum_{n=1}^{N} \lambda_0(I_n) = \lambda_0(\dot\bigcup_{n=1}^{N} I_n) \leqslant \lambda_0(\dot\bigcup_{n\in\mathbb{N}} I_n)$$

$$\sum_{n=1}^{\infty} \lambda_0(I_n) = \sup\{N|\sum_{n=1}^{N} \lambda_0(I_n)\} \leqslant \lambda_0(\dot\bigcup_{n\in\mathbb{N}} I_n)$$

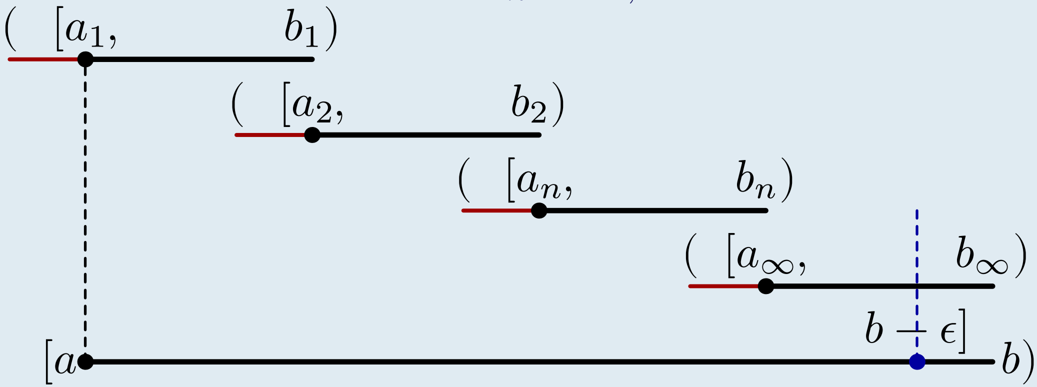
NOTE: we have used increasing and finite additivity of $\lambda_0$ (need a proof).

# Countable Additivity of $(\mathbb{R}, \mathcal{S}, \lambda_0)$ (2)

Proof of subgoal 2 (hard):

- For $0 < \epsilon$, let $I^o_{n,\epsilon} = (a_n - 2^{-n}\epsilon, b_n)$, thus $\lambda_0(I^o_{n,\epsilon}) = \lambda_0(I_n) + 2^{-n}\epsilon$
- By Heine-Borel theorem there exists $N$ such that

  $[a, b-\epsilon) \subset [a, b-\epsilon] \subseteq \dot{\bigcup}_{n=1}^{N} I^o_{n,\epsilon}$, thus $\lambda_0([a, b-\epsilon)) \leqslant \sum_{n=1}^{N} \lambda_0(I^o_{n,\epsilon})$




- $\lambda_0([a, b)) = \lambda_0([a, b-\epsilon)) + \epsilon \leqslant \sum_{n=1}^{N} \lambda_0(I^o_{n,\epsilon}) + \epsilon$
  $\leqslant \sum_{n=1}^{N} \lambda_0(I_n) + \sum_{n=1}^{N} 2^{-n}\epsilon + \epsilon$
  $\leqslant \sum_{n=1}^{N} \lambda_0(I_n) + 2\epsilon \leqslant \sum_{n=1}^{\infty} \lambda_0(I_n) + 2\epsilon.$

# Measurable Mappings

Let $(X, \mathcal{A})$ and $(X', \mathcal{A}')$ be measurable spaces. A map $T : X \to X'$ is called $\mathcal{A}/\mathcal{A}'$-*measurable* (or *meausurable*) if the pre-image (aka inverse mapping) of every measurable set in $\mathcal{A}'$ is a measurable set in $\mathcal{A}$:

$$T^{-1}(A') \cap X \in \mathcal{A} \qquad \forall A' \in \mathcal{A}'$$

where $T^{-1}(A') := \{x \,|\, T(x) \in \mathcal{A}'\}$ (and $T^{-1}(\mathcal{A}') := \{T^{-1}(A') \,|\, A' \in \mathcal{A}'\}$).

```
[pred_setTheory.PREIMAGE_def]
 ⊢ PREIMAGE f s = {x | f x ∈ s}

[sigma_algebraTheory.measurable_def]
 ⊢ measurable a b =
   {f |
    f ∈ (space a → space b) ∧
    ∀s. s ∈ subsets b ⇒
        PREIMAGE f s ∩ space a ∈ subsets a}
```

NOTE: For $T : X \to X'$, there may exist $x$ such that $x \notin X$ but $T(x) \in X'$.

# Borel Measurable Functions

A *measurable function* is a measurable map $u : X \to \mathbb{R}$ (or $\overline{\mathbb{R}}$) from a measurable space $(X, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$ (or $(\overline{\mathcal{B}}, \overline{\mathbb{R}})$). Thus,

$$u^{-1}(B) \cap X \in \mathcal{F} \qquad \forall B \in \mathcal{B}$$

In particular, a measurable map $\xi : \Omega \to \overline{\mathbb{R}}$ from a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ (i.e., $P(\Omega) = 1$) to $(\overline{\mathcal{B}}, \overline{\mathbb{R}})$ is called a *random variable* (r.v.).

**[Notations]** Let $\xi$ be a r.v. and $E = \{\xi \leqslant 5\} \in \mathcal{F}$ be an event, we have, e.g.

$$\mathrm{P}\{\xi \leqslant 5\} := \mathrm{P}(\{\xi \leqslant 5\}) := \mathrm{P}(\{\omega \in \Omega \,|\, \xi(\omega) \leqslant 5\})$$

In general, given $(\Omega, \mathcal{F}, \mathrm{P})$, for any $B \in \overline{\mathcal{B}}$ we have $\{\xi \in B\} \in \mathcal{F}$ and

$$\mathrm{P}\{\xi \in B\} := \mathrm{P}(\{\omega \in \Omega \,|\, \xi(\omega) \in B\}) = (\mathrm{P} \circ \xi^{-1})\,(B)$$

And $(\overline{\mathbb{R}}, \overline{\mathcal{B}}, \mathrm{P} \circ \xi^{-1})$ forms another probability measure space.

# Supplement: Random Variables

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be the probability space. $X$ and $Y$ are r.v.'s.

- $X + Y$ is abbreviation of $\lambda\omega.\, X(\omega) + Y(\omega)\ :\ \Omega \to \mathbb{R}$ or $\overline{\mathbb{R}}$.
- $X^2$ is abbreviation of $\lambda\omega.\, X(\omega)^2$.

In the "continuous" elementary probability, the probability space is $([0,1], \mathcal{F}, \mathrm{P})$ where $\mathcal{F} = \mathcal{B} \cap [0,1]$, and

- A random variable is a measurable mapping from $[0,1]$ to $\mathbb{R}$, e.g. $X(x) = x$ or $X(x) = \sin x$.
- The expectation (and variance, etc.) of r.v.'s can be calculated by Riemann integration.

# Proving Borel Measurable Functions (1)

It's hard to prove a function measurable by checking $\forall B \in \mathcal{B}$.

Some alternative definitions:

```
[real_borelTheory.in_borel_measurable_open]
 ⊢ f ∈ borel_measurable M  ⟺
   ∀s. s ∈ subsets (sigma 𝒰(:real) {s | open s}) ⇒
       PREIMAGE f s ∩ space M ∈ subsets M

[real_borelTheory.in_borel_measurable_gr]
 ⊢ sigma_algebra m ⇒
   (f ∈ borel_measurable m ⟺
    f ∈ (space m → 𝒰(:real)) ∧
    ∀a. {w | w ∈ space m ∧ a < f w} ∈ subsets m)

[real_borelTheory.in_borel_measurable_const]
 ⊢ sigma_algebra a ∧ (∀x. x ∈ space a ⇒ f x = k) ⇒
   f ∈ borel_measurable a

[real_borelTheory.in_borel_measurable_continuous_on]
 ⊢ f continuous_on 𝒰(:real) ⇒
   f ∈ borel_measurable borel
```

# Proving Borel Measurable Functions (2)

Alternative definitions of extended Borel measurable functions:

[borelTheory.IN_MEASURABLE_BOREL_ALT1]
$\vdash$ sigma_algebra $a \Rightarrow$
  ($f \in$ Borel_measurable $a \iff$
   $f \in$ (space $a \to \mathcal{U}$(:extreal)) $\wedge$
   $\forall c$. $\{x$ | Normal $c \leq f\ x\} \cap$ space $a \in$ subsets $a$)

[borelTheory.IN_MEASURABLE_BOREL_ALT4]
$\vdash$ sigma_algebra $a \wedge$ ($\forall x$. $x \in$ space $a \Rightarrow f\ x \neq -\infty$) $\Rightarrow$
  ($f \in$ Borel_measurable $a \iff$
   $f \in$ (space $a \to \mathcal{U}$(:extreal)) $\wedge$
   $\forall c\ d$.
     $\{x$ | Normal $c \leq f\ x \wedge f\ x <$ Normal $d\} \cap$
     space $a \in$ subsets $a$)

[borelTheory.IN_MEASURABLE_BOREL_IMP_BOREL]
$\vdash f \in$ borel_measurable (measurable_space $m$) $\Rightarrow$
  Normal $\circ f \in$
  Borel_measurable (measurable_space $m$)

# Proving Borel Measurable Functions (3)

Arithmetic compositions of Borel measurable functions are still measurable:

$\vdash$ sigma_algebra $a$ $\wedge$ $f$ $\in$ borel_measurable $a$ $\wedge$
$g$ $\in$ borel_measurable $a$ $\wedge$
($\forall x$. $x$ $\in$ space $a$ $\Rightarrow$ $h$ $x$ = $f$ $x$ $+$ $g$ $x$) $\Rightarrow$
$h$ $\in$ borel_measurable $a$

$\vdash$ sigma_algebra $a$ $\wedge$ $f$ $\in$ borel_measurable $a$ $\wedge$
$g$ $\in$ borel_measurable $a$ $\wedge$
($\forall x$. $x$ $\in$ space $a$ $\Rightarrow$ $h$ $x$ = $f$ $x$ $\times$ $g$ $x$) $\Rightarrow$
$h$ $\in$ borel_measurable $a$

$\vdash$ sigma_algebra $a$ $\wedge$ $f$ $\in$ borel_measurable $a$ $\wedge$
($\forall x$. $x$ $\in$ space $a$ $\Rightarrow$ $g$ $x$ = $z$ $\times$ $f$ $x$) $\Rightarrow$
$g$ $\in$ borel_measurable $a$

$\vdash$ sigma_algebra $a$ $\wedge$ $f$ $\in$ borel_measurable $a$ $\wedge$
$g$ $\in$ borel_measurable $a$ $\Rightarrow$
($\lambda x$. max ($f$ $x$) ($g$ $x$)) $\in$ borel_measurable $a$

$\vdash$ sigma_algebra $a$ $\wedge$ ($\forall x$. $x$ $\in$ space $a$ $\Rightarrow$ $f$ $x$ = $k$) $\Rightarrow$
$f$ $\in$ borel_measurable $a$

$\vdash$ sigma_algebra $a$ $\wedge$ $f$ $\in$ borel_measurable $a$ $\wedge$
$s$ $\in$ subsets $a$ $\Rightarrow$
($\lambda x$. $f$ $x$ $\times$ indicator $s$ $x$) $\in$ borel_measurable $a$

# Proving Borel Measurable Functions (4)

**Proposition**: Let $f, g$ be Borel measurable functions from $(X, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$, show that $f + g$ is Borel measurable.

**Proof** (by showing $\{x \mid x \in X \wedge f(x) + g(x) < c\} \in \mathcal{A}$ for all $c \in \mathbb{R}$)

1. $\{x \mid x \in X \wedge f(x) + g(x) < c\} = \{x \mid x \in X \wedge f(x) < c - g(x)\}$
2. $\{x \mid x \in X \wedge f(x) < c - g(x)\} = \bigcup_{r \in \mathbb{Q}} \{x \mid x \in X \wedge f(x) < r \wedge r < c - g(x)\}$
3. it suffices to show $\{x \mid x \in X \wedge f(x) < r \wedge r < c - g(x)\} \in \mathcal{A}$ for any $r$
4. $\{x \mid x \in X \wedge f(x) < r \wedge r < c - g(x)\} = \{x \mid x \in X \wedge f(x) < r\} \cap \{x \mid x \in X \wedge r < c - g(x)\}$
5. $\{x \mid x \in X \wedge f(x) < r\} \in \mathcal{A}$ by the same alternative definition.
6. $\{x \mid x \in X \wedge r < c - g(x)\} = \{x \mid x \in X \wedge g(x) < c + r\}$ (same as above).

NOTE: (2) has used the $\mathbb{Q}$-dense property: $\forall x, y \in \mathbb{R}. \exists r \in \mathbb{Q}. x < r \wedge r < y$.

# The need of Lebesgue integration

```
[probabilityTheory.expectation_def]
 ⊢ expectation = ∫

[probabilityTheory.variance_alt]
 ⊢ variance p X =
   expectation p (λ x. (X x − expectation p X)²)
```

Converting the usual Riemann integration to Lebesgue integration:

$$\int_a^b f(x)\,\mathrm{d}x = \int_m \lambda x.\, f(x) \cdot \mathbb{1}_s(x), \qquad m = (\mathbb{R}, \mathcal{B}, \lambda),\, s = [a,b],\, f\colon \mathbb{R} \to \overline{\mathbb{R}}$$

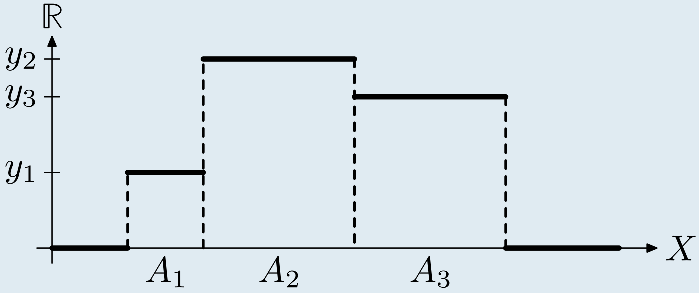
Steps to establish Lebesgue integration:

- Positive simple functions;
- Integration of simple functions;
- Integration of positive functions;
- Integration of measurable functions.

# Positive simple functions $\mathcal{E}^+(\mathcal{A})$

A *positive simple function* $f\colon X \to \mathbb{R}$ on $(X, \mathcal{A})$ is a function of the form

$$f(x) = \sum_{m=1}^{M} y_m \mathbb{1}_{A_m}(x), \qquad M \in \mathbb{N},\ y_m \in \mathbb{R}^+,\ A_m \in \mathcal{A} \text{ disjoint}$$




The *standard representation* of $f$ is the following form:

$$f(x) = \sum_{n=0}^{N} z_n \mathbb{1}_{B_n}(x), \qquad M \in \mathbb{N},\ z_n \in \mathbb{R}^+,\ B_n \in \mathcal{A} \text{ disjoint},\ X = \bigcup_{n=0}^{N} B_n$$

(This can be done by setting $B_0 = X \setminus \bigcup_{n=1}^{N} B_n$ and $z_0 = 0$.)

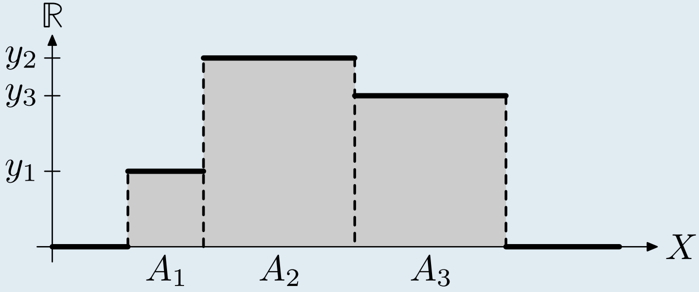
# Integration of positive simple functions

Let $(X, \mathcal{A}, \mu)$ be a measure space, and $f \in \mathcal{E}^+$ a positive simple function

$$f(x) = \sum_{i=0}^{M} y_i \mathbb{1}_{A_i}(x), \quad M \in \mathbb{N}, y_i \in \mathbb{R}^+, A_i \in \mathcal{A} \text{ disjoint}, X = \bigcup_{i=0}^{M} A_i$$

The *($\mu$-)integral* of $f$ us the $\mu$-area enclosed by the graph of $f$ and the abscissa ($X$-axis), which is independent of the representation of $f$:

$$I_\mu(f) := \sum_{i=0}^{M} y_i \mu(A_i) \in [0, \infty]$$



# Integration of positive functions

Let $(X, \mathcal{A}, \mu)$ be a measure space, the *($\mu$-)integral* of a positive function $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}$ is given by

$$\int_{(X,\mathcal{A},\mu)} u \text{ or } \int u \,\mathrm{d}\mu := \sup\{I_\mu(g) \mid g \leqslant u, g \in \mathcal{E}^+(\mathcal{A})\} \in [0, \infty].$$

Given any $u \in \mathcal{M}^+_{\overline{\mathbb{R}}}$, it's possible to construct a sequence of positive simple functions $f_n \in \mathcal{E}^+(\mathcal{A})$, $n \in \mathbb{N}$ such that

- $\forall x \in X.\, 0 \leqslant f_n(x) \leqslant f_{n+1}(x) \leqslant u(x)$;
- $\forall x \in X.\, \sup\{f_n(x)\} = u(x)$.

Beppo Levi (Monotone Convergence): $u, u_i \in \mathcal{M}^+_{\overline{\mathbb{R}}}$ and $u = \sup_{n\in\mathbb{N}} u_n$, then

$$\int \sup_{n\in\mathbb{N}} u_n \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int u_n \,\mathrm{d}\mu$$

# Measurable functions and positive functions

Let $(X, \mathcal{A}, \mu)$ be a measure space, any measurable function $u \in \mathcal{M}_{\overline{\mathbb{R}}}$ can be decomposed into two positive functions $u^+, u^- \in \mathcal{M}^+_{\overline{\mathbb{R}}}$:

- $u^+(x) := \max\{u(x), 0\}$;
- $u^-(x) := -\min\{u(x), 0\}$;
- $u = u^+ - u^-$;
- $|u| = u^+ + u^-$.

# Integration of measurable functions

A function $u\colon X \to \overline{\mathbb{R}}$ on a measure space $(X, \mathcal{A}, \mu)$ is said to be *($\mu$-)integrable*, if it's $\mathcal{A}/\overline{\mathbb{R}}$-measurable and if

$$\int u^+ \,\mathrm{d}\mu < \infty, \qquad \int u^- \,\mathrm{d}\mu < \infty$$

In this case, the *($\mu$-)integral* of $u$ is

$$\int u \,\mathrm{d}\mu := \int u^+ \,\mathrm{d}\mu - \int u^- \,\mathrm{d}\mu \in (-\infty, \infty)\,.$$

We write $\mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$ for the set of all $\overline{\mathbb{R}}$-valued $\mu$-integrable functions.

NOTE: If just one of $\int u^+ \,\mathrm{d}\mu$ and $\int u^- \,\mathrm{d}\mu$ is infinite,
$u$ is not integrable but the integration $\int u \,\mathrm{d}\mu$ is still well-defined by the above formula.

# Related formal definitions in HOL4 (1)

```
[lebesgueTheory.pos_simple_fn_def]
⊢ pos_simple_fn m f s a x ⇔
    (∀t. t ∈ m_space m ⇒ 0 ≤ f t) ∧
    (∀t. t ∈ m_space m ⇒
         f t = ∑ (λi. Normal (x i) × 𝟙 (a i) t) s) ∧
    (∀i. i ∈ s ⇒ a i ∈ measurable_sets m) ∧
    FINITE s ∧ (∀i. i ∈ s ⇒ 0 ≤ x i) ∧
    (∀i j.
       i ∈ s ∧ j ∈ s ∧ i ≠ j ⇒ DISJOINT (a i) (a j)) ∧
    ⋃ (IMAGE a s) = m_space m

[lebesgueTheory.pos_simple_fn_integral_def]
⊢ pos_simple_fn_integral m s a x =
    ∑ (λi. Normal (x i) × measure m (a i)) s

[lebesgueTheory.psfs_def]
⊢ psfs m f = {(s,a,x) | pos_simple_fn m f s a x}

[lebesgueTheory.psfis_def]
⊢ psfis m f =
    IMAGE
      (λ(s,a,x). pos_simple_fn_integral m s a x)
      (psfs m f)
```

# Related formal definitions in HOL4 (2)

[lebesgueTheory.pos_fn_integral_def]

$$\vdash \int^{+} m\ f = \texttt{sup}\ \{r \mid (\exists g.\ r \in \texttt{psfis}\ m\ g \land \forall x.\ x \in \texttt{m\_space}\ m \Rightarrow g\ x \le f\ x)\}$$

$$\vdash \texttt{integrable}\ m\ f \iff f \in \texttt{Borel\_measurable}\ (\texttt{measurable\_space}\ m) \land \int^{+} m\ f^{+} \neq +\infty \land \int^{+} m\ f^{-} \neq +\infty$$ [integrable_def]

$$\vdash \int m\ f = \int^{+} m\ f^{+} - \int^{+} m\ f^{-}$$ [integral_def]

[lebesgueTheory.lebesgue_monotone_convergence (Beppo Levi)]

$$\begin{aligned}
\vdash\ & \texttt{measure_space}\ m \land \\
& (\forall i.\ fi\ i \in \texttt{Borel_measurable}\ (\texttt{measurable_space}\ m)) \land \\
& (\forall i\ x.\ x \in \texttt{m_space}\ m \Rightarrow 0 \le fi\ i\ x) \land \\
& (\forall x.\ x \in \texttt{m_space}\ m \Rightarrow \texttt{mono_increasing}\ (\lambda i.\ fi\ i\ x)) \land \\
& (\forall x.\ x \in \texttt{m_space}\ m \Rightarrow \texttt{sup}\ (\texttt{IMAGE}\ (\lambda i.\ fi\ i\ x)\ \mathcal{U}(:\texttt{num})) = f\ x) \Rightarrow \\
& \int^{+} m\ f = \texttt{sup}\ (\texttt{IMAGE}\ (\lambda i.\ \int^{+} m\ (fi\ i))\ \mathcal{U}(:\texttt{num}))
\end{aligned}$$

# References

Chapter 7–10 of

R. L. Schilling, Measures, Integrals and Martingales (2nd Edition). Cambridge University Press, 2017.