HOL Theorem Proving and Formal Probability (5)

Chun TIAN
chun.tian@anu.edu.au

17/04/2024

Properties of Lebesgue Integrals (1)

Let $(X, \mathscr{A}, \mu)$ be a measure space and $u, v \in \mathcal{L}^1_{\mathbb{R}}(\mu)$, $\alpha \in \mathbb{R}$, then

$$\alpha u \in \mathcal{L}^1_{\mathbb{R}}(\mu) \text{ and } \int \alpha u \,\mathrm{d}\mu = \alpha \int u \,\mathrm{d}\mu, \tag{1.a}$$

$$u + v \in \mathcal{L}^1_{\mathbb{R}}(\mu) \text{ and } \int (u+v) \,\mathrm{d}\mu = \int u \,\mathrm{d}\mu + \int v \,\mathrm{d}\mu, \tag{1.b}$$

$$\min\{u, v\}, \max\{u, v\} \in \mathcal{L}^1_{\mathbb{R}}(\mu), \tag{1.c}$$

$$u \leqslant v \Longrightarrow \int u \,\mathrm{d}\mu \leqslant \int v \,\mathrm{d}\mu, \tag{1.d}$$

$$\left| \int u \,\mathrm{d}\mu \right| \leqslant \int |u| \,\mathrm{d}\mu. \tag{1.e}$$

Proof of (1.e): using $\pm u \leqslant |u|$,

$$\left| \int u \,\mathrm{d}\mu \right| = \max\left\{ \int u \mathrm{d}\mu, -\int u \mathrm{d}\mu \right\} \leqslant \max\left\{ \int |u| \mathrm{d}\mu, \int |-u| \mathrm{d}\mu \right\} = \int |u| \mathrm{d}\mu$$

Properties of Lebesgue Integrals (2)

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \Rightarrow$$
$$\texttt{integrable}\ m\ (\lambda x.\ \texttt{Normal}\ c \times f\ x) \qquad \texttt{[integrable\_cmul]}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \Rightarrow$$
$$\int m\ (\lambda x.\ \texttt{Normal}\ c \times f\ x) = \texttt{Normal}\ c \times \int m\ f \qquad \texttt{[integral\_cmul]}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \wedge$$
$$\texttt{integrable}\ m\ g \Rightarrow$$
$$\texttt{integrable}\ m\ (\lambda x.\ f\ x + g\ x) \qquad \texttt{[integrable\_add']}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \wedge$$
$$\texttt{integrable}\ m\ g \Rightarrow$$
$$\int m\ (\lambda x.\ f\ x + g\ x) = \int m\ f + \int m\ g \qquad \texttt{[integral\_add']}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f_1 \wedge$$
$$\texttt{integrable}\ m\ f_2 \wedge$$
$$(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow f_1\ x \leq f_2\ x) \Rightarrow$$
$$\int m\ f_1 \leq \int m\ f_2 \qquad \texttt{[integral\_mono]}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \Rightarrow$$
$$\texttt{integrable}\ m\ (\texttt{abs} \circ f) \qquad \texttt{[integrable\_abs]}$$

$$\vdash \texttt{measure\_space}\ m \wedge \texttt{integrable}\ m\ f \Rightarrow$$
$$\texttt{abs}\ (\int m\ f) \leq \int^{+} m\ (\texttt{abs} \circ f) \qquad \texttt{[integral\_triangle\_ineq']}$$

Properties of Lebesgue Integrals (3)

$\vdash$ $\texttt{measure\_space}\ m\ \wedge$
$(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow 0 \le f\ x) \wedge 0 \le c \Rightarrow$
$\int^{+} m\ (\lambda x.\ \texttt{Normal}\ c \times f\ x) = \texttt{Normal}\ c \times \int^{+} m\ f$

$\vdash$ $\texttt{measure\_space}\ m\ \wedge$
$(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow 0 \le f\ x) \wedge$
$(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow 0 \le g\ x) \wedge$
$f \in \texttt{Borel\_measurable}\ (\texttt{measurable\_space}\ m) \wedge$
$g \in \texttt{Borel\_measurable}\ (\texttt{measurable\_space}\ m) \Rightarrow$
$\int^{+} m\ (\lambda x.\ f\ x + g\ x) = \int^{+} m\ f + \int^{+} m\ g$

$\vdash$ $(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow 0 \le f\ x) \wedge$
$(\forall x.\ x \in \texttt{m\_space}\ m \Rightarrow f\ x \le g\ x) \Rightarrow$
$\int^{+} m\ f \le \int^{+} m\ g$

$\vdash$ $\texttt{measure\_space}\ m \wedge 0 \le z\ \wedge$
$\texttt{pos\_simple\_fn}\ m\ f\ s\ a\ x \Rightarrow$
$\exists s'\ a'\ x'.$
$\quad\texttt{pos\_simple\_fn}\ m\ (\lambda t.\ \texttt{Normal}\ z \times f\ t)\ s'\ a'\ x'\ \wedge$
$\quad\texttt{pos\_simple\_fn\_integral}\ m\ s'\ a'\ x' =$
$\quad\texttt{Normal}\ z \times \texttt{pos\_simple\_fn\_integral}\ m\ s\ a\ x$

$\vdash$ $\texttt{measure\_space}\ m \wedge \texttt{pos\_simple\_fn}\ m\ f\ s\ a\ x\ \wedge$
$\texttt{pos\_simple\_fn}\ m\ g\ s\ a\ y \Rightarrow$
$\texttt{pos\_simple\_fn\_integral}\ m\ s\ a\ x\ +$
$\texttt{pos\_simple\_fn\_integral}\ m\ s\ a\ y =$
$\texttt{pos\_simple\_fn\_integral}\ m\ s\ a\ (\lambda i.\ x\ i + y\ i)$

Null Sets

Let $(X, \mathcal{A}, \mu)$ be a measure space, A (μ-)*null set* $N \in \mathcal{N}_\mu \subseteq \mathcal{A}$ is a measurable set $N \in \mathcal{A}$ satisfying

$$N \in \mathcal{N}_\mu \Longleftrightarrow N \in \mathcal{A} \ \wedge \ \mu(N) = 0$$

Null sets are closed under countable (and finite) union and intersection:

$$N_1, N_2, \ldots \in \mathcal{N}_\mu \Longrightarrow \bigcap_{i \in \mathbb{N}} N_i, \bigcup_{i \in \mathbb{N}} N_i \in \mathcal{N}_\mu$$

Attention: $N \in \mathcal{N}_\mu$ and $N' \subseteq N$ does NOT imply $N' \in \mathcal{N}_\mu$.

Complete measure space:

$\vdash$ $\texttt{complete\_measure\_space}\ m \iff$
$\quad \texttt{measure\_space}\ m\ \wedge$
$\quad \forall s.\ \texttt{null\_set}\ m\ s \Rightarrow$
$\qquad \forall t.\ t \subseteq s \Rightarrow t \in \texttt{measurable\_sets}\ m$

The 'Almost Everywhere'

If a property $P(x)$ holds for all $x \in X$ except for a null set $N \in \mathcal{N}_\mu$, i.e.

$$\exists N \in \mathcal{N}_\mu.\, \forall x \in X \setminus N.\, P(x)$$

We say that $P(x)$ holds (μ-)*almost everywhere (a.e.)*.

$\vdash$ (AE x::m. P x) $\iff$
  $\exists N$. null_set m N $\wedge$
    $\forall x$. $x \in$ m_space m DIFF N $\Rightarrow$ P x [borelTheory.AE_DEF]

Let $u, v \in \mathcal{L}^1_{\mathbb{R}}(\mu)$, u is a.e. equal to v (or $u = v$ (μ-)a.e.), if

$$\{x \in X \,|\, u(x) \neq v(x)\} \text{ is (contained in) a } \mu\text{-null set } N \in \mathcal{N}_\mu \subseteq \mathcal{A}$$

In this case, we have

$$\int u \,\mathrm{d}\mu = \int_{N^c} u \,\mathrm{d}\mu + \int_N u \,\mathrm{d}\mu = \int_{N^c} u \,\mathrm{d}\mu + 0 = \int_{N^c} v \,\mathrm{d}\mu + \int_N v \,\mathrm{d}\mu = \int v \,\mathrm{d}\mu$$

The 'Almost Everywhere' (2)

Let $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathcal{A})$ be a measurable function on a measure space $(X, \mathcal{A}, \mu)$. Then

$$\int |u| \,\mathrm{d}\mu = 0 \iff |u| = 0 \text{ a.e.} \iff \mu\{u \neq 0\} = 0 \qquad (2.\mathrm{a})$$

$$\mathbb{1}_N u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu) \text{ for all } N \in \mathcal{N}_\mu \text{ and } \int_N u \,\mathrm{d}\mu = 0 \qquad (2.\mathrm{b})$$

Markov inequality (for measure space): For all $u \in \mathcal{L}^1_{\overline{\mathbb{R}}}(\mu)$, $A \in \mathcal{A}$ and $c > 0$:

$$\mu(\{|u| \geqslant c\} \cap A) \leqslant \frac{1}{c} \int_A |u| \,\mathrm{d}\mu$$

$$\begin{aligned} \mu(\{|u| \geqslant c\} \cap A) &= \int \mathbb{1}_{\{|u| \geqslant c\} \cap A}(x) \,\mathrm{d}\mu = \int_A \frac{c}{c} \mathbb{1}_{\{|u| \geqslant c\}}(x) \,\mathrm{d}\mu \\ &\leqslant \frac{1}{c} \int_A |u(x)| \, \mathbb{1}_{\{|u| \geqslant c\}}(x) \,\mathrm{d}\mu \leqslant \frac{1}{c} \int_A |u(x)| \,\mathrm{d}\mu \end{aligned}$$

Application of 'Almost Everywhere'

Dirichlet function $D(x)\colon \mathbb{R} \to \mathbb{R}$ (not Riemann integrable):

- $x \in \mathbb{Q} \Longrightarrow D(x) = 1$
- $x \notin \mathbb{Q} \Longrightarrow D(x) = 0$

Since any countable set (including $\mathbb{Q}$) is also a null set, we have

$$\int D(x)\,\mathrm{d}\mu = \int_{\mathbb{Q}} D(x)\,\mathrm{d}\mu + \int_{\mathbb{R}\setminus\mathbb{Q}} D(x)\,\mathrm{d}\mu = \int_{\mathbb{Q}} 1\mathrm{d}\mu + \int_{\mathbb{R}\setminus\mathbb{Q}} 0\,\mathrm{d}\mu = 0 + 0 = 0$$

Thus Dirichlet function $D(x)$ is Lebesgue integrable (and the integral is zero).

TODO: This theorem is not yet in HOL4.

Density Measure Space

On the measure space $(X, \mathcal{A}, \mu)$, let $u \in \mathcal{M}^+_{\mathbb{R}}(\mathcal{A})$, the set function

$$\nu : A \mapsto \int_A u \, \mathrm{d}\mu = \int \mathbb{1}_A u \, \mathrm{d}\mu, \qquad A \in \mathcal{A}$$

is a measure on $(X, \mathcal{A})$. It's called the measure with density (function) u w.r.t. μ. We write $\nu = u \cdot \mu$ (or $\mathrm{d}\nu = u\mathrm{d}\mu$).

$\vdash f \times m$ = $(\lambda s.\ \int^+ m\ (\lambda x.\ f\ x \times \mathbb{1}\ s\ x))$ [density_measure_def]
$\vdash$ density $m\ f$ =
(m_space m,measurable_sets m,$f \times m$) [density_def]

$\vdash$ measure_space m $\wedge$
$f \in$ Borel_measurable (measurable_space m) $\wedge$
$(\forall x.\ x \in$ m_space $m \Rightarrow 0 \leq f\ x) \Rightarrow$
measure_space (density $m\ f$) [measure_space_density]

Integration of Density Measure

Given $\nu = u \cdot \mu$ (or $\mathrm{d}\nu = u\mathrm{d}\mu$),

$$\int g(x)\,\mathrm{d}\nu = \int g(x)\,u(x)\,\mathrm{d}\mu$$

$\vdash$ $f \times m$ = $(\lambda s.\ \int^{+} m\ (\lambda x.\ f\ x \times \mathbb{1}\ s\ x))$ [density_measure_def]
$\vdash$ density $m\ f$ =
(m_space m,measurable_sets m,$f \times m$) [density_def]

[lebesgueTheory.pos_fn_integral_density]
$\vdash$ measure_space m $\wedge$
$f \in$ Borel_measurable (measurable_space m) $\wedge$
$g \in$ Borel_measurable (measurable_space m) $\wedge$
(AE x::m. $0 \le f\ x$) $\wedge$ $(\forall x.\ 0 \le g\ x)$ $\Rightarrow$
$\int^{+}$ (density m f^{+}) g = $\int^{+} m\ (\lambda x.\ f^{+}\ x \times g\ x)$

TODO: Improve $\forall x. 0 \leqslant g(x)$ to $\forall x \in X. 0 \leqslant g(x)$ in [pos_fn_integral_density].

Distribution Measure Space

On the measure space $(X, \mathcal{A}, \mu)$, let $u \in \mathcal{M}_{\overline{\mathbb{R}}}(\mathcal{A})$, the set function

$$\nu : A \mapsto \mu(u^{-1}(A) \cap X)$$

is called the distribution measure of u (under μ). $(\overline{\mathbb{R}}, \overline{\mathcal{B}}, \nu)$ is measure space.

$\vdash$ distr m f =
  ($\lambda\, s$. measure m (PREIMAGE f s $\cap$ m_space m))   [distr_def]

[measure_space_distr]
$\vdash$ measure_space M $\wedge$ sigma_algebra B $\wedge$
  f $\in$ measurable (measurable_space M) B $\Rightarrow$
  measure_space (space B,subsets B,distr M f)

In probability space, a *distribution function* is a mono-increasing function from $\overline{\mathbb{R}}$ to $[0, 1]$:

$\vdash$ distribution_function p X =
  ($\lambda\, x$. prob p ($\{w$ | X w $\leq$ $x\}$ $\cap$ p_space p))

Integration of Distribution Measure

Write $T(\mu)(A) := \mu(T^{-1}(A))$, the following *transformation theorem* holds

$$\int u \, \mathrm{d}T(\mu) = \int u \circ T \, \mathrm{d}\mu$$

```
⊢ measure_space M ∧ sigma_algebra B ∧
  f ∈ measurable (measurable_space M) B ∧
  u ∈ Borel_measurable B ⇒
  ∫ (space B,subsets B,distr M f) u = ∫ M (u ∘ f) ∧
  (integrable (space B,subsets B,distr M f) u ⇔
   integrable M (u ∘ f))          [martingaleTheory.integral_distr]
```

The Radon-Nikodým Theorem

Let μ, ν be two measures on the measurable space $(X, \mathcal{A})$. We call v *absolutely continuous* w.r.t. μ and write $\nu \ll \mu$ if

$$N \in \mathcal{A}, \mu(N) = 0 \Longrightarrow \nu(N) = 0$$

In other words, all μ-null sets are ν-null sets: $\mathcal{N}_\mu \subseteq \mathcal{N}_\nu$.

The Radon-Nikodým Theorem has the position of Fundamental Theorem of Calculus in Lebesgue Integration:

If μ is σ-finite, then the following assertaions are equivalent:

- $\nu(A) = \int_A f(x)\, \mu(\mathrm{d}x)$ for some a.e. unique $f \in \mathcal{M}^+(\mathcal{A})$;
- $\nu \ll \mu$.

The unique function f is called the *Radon-Nikodým derivative* and is traditionally denoted by $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

The Radon-Nikodým Theorem in HOL4

```
[lebesgueTheory.measure_absolutely_continuous_def]
⊢ v ≪ m ⟺
    ∀s. s ∈ measurable_sets m ∧ measure m s = 0 ⇒
        v s = 0

[lebesgueTheory.Radon_Nikodym']
⊢ measure_space m ∧ sigma_finite m ∧
    measure_space (m_space m,measurable_sets m,v) ∧
    v ≪ m ⇒
    ∃f. f ∈ Borel_measurable (measurable_space m) ∧
        (∀x. x ∈ m_space m ⇒ 0 ≤ f x) ∧
        ∀s. s ∈ measurable_sets m ⇒ (f × m) s = v s

[lebesgueTheory.RN_deriv_def]
⊢ v / m =
    εf. f ∈ Borel_measurable (measurable_space m) ∧
        (∀x. x ∈ m_space m ⇒ 0 ≤ f x) ∧
        ∀s. s ∈ measurable_sets m ⇒ (f × m) s = v s
```

Application of The Radon-Nikodým Theorem

```
[probabilityTheory.prob_density_function_def (pdf_def)]
⊢ prob_density_function p X =
  distribution p X / ext_lborel

[probabilityTheory.pdf_le_pos]
⊢ prob_space p ∧ random_variable X p Borel ∧
  distribution p X ≪ ext_lborel ⇒
  0 ≤ prob_density_function p X x

[probabilityTheory.expectation_pdf_1]
⊢ prob_space p ∧ random_variable X p Borel ∧
  distribution p X ≪ ext_lborel ⇒
  expectation ext_lborel
    (prob_density_function p X) =
  1
```

PDF (probability density function) = *derivative* of distribution function

```
[probabilityTheory.distribution_function_def]
⊢ distribution_function p X =
  (λx. prob p ({w | X w ≤ x} ∩ p_space p))
```