# HOL Theorem Proving and Formal Probability (6)

Chun TIAN
chun.tian@anu.edu.au

01/05/2024

# Probability Theory: Basic Definitions

$\vdash$ `prob_space` $p$ $\iff$
`measure_space` $p$ $\wedge$ `measure` $p$ (`m_space` $p$) = 1
$\vdash$ `p_space = m_space`
$\vdash$ `events = measurable_sets`
$\vdash$ `prob = measure`

$\vdash$ `random_variable` $X$ $p$ $s$ $\iff$
$X$ $\in$ `measurable` (`p_space` $p$,`events` $p$) $s$
$\vdash$ `real_random_variable` $X$ $p$ $\iff$
`random_variable` $X$ $p$ `Borel` $\wedge$
$\forall x.$ $x \in$ `p_space` $p$ $\Rightarrow$ $X$ $x \neq -\infty$ $\wedge$ $X$ $x \neq +\infty$

$\vdash$ `expectation` = $\int$
$\vdash$ `prob_space` $p$ $\wedge$ `FINITE` (`p_space` $p$) $\wedge$
`real_random_variable` $X$ $p$ $\wedge$ `integrable` $p$ $X$ $\Rightarrow$
`expectation` $p$ $X$ =
$\sum$ ($\lambda r.$ $r$ $\times$ `prob` $p$ (`PREIMAGE` $X$ $\{r\}$ $\cap$ `p_space` $p$))
(`IMAGE` $X$ (`p_space` $p$)) [finite_expectation1]

$\vdash$ `distribution` $p$ $X$ =
($\lambda s.$ `prob` $p$ (`PREIMAGE` $X$ $s$ $\cap$ `p_space` $p$)) [distribution_def]
$\vdash$ `distribution = distr` [distribution_distr]

# (Second) Moment and Variance

$\mathrm{E}[X^2]$ or $\mathrm{E}[(X-c)^2]$ or $\mathrm{D}[X] := \mathrm{E}[(X - \mathrm{E}[X])^2]$ plays an important role.

$\vdash$ `moment` $p\ X\ a\ r$ =
`expectation` $p$ ($\lambda x$. ($X\ x - a$) `pow` $r$)

$\vdash$ `absolute_moment` $p\ X\ a\ r$ =
`expectation` $p$ ($\lambda x$. `abs` ($X\ x - a$) `pow` $r$)

$\vdash$ `central_moment` $p\ X\ r$ =
`moment` $p\ X$ (`expectation` $p\ X$) $r$

$\vdash$ `second_moment` $p\ X\ a$ = `moment` $p\ X\ a\ 2$

$\vdash$ `variance` $p\ X$ = `central_moment` $p\ X\ 2$

$\vdash$ `standard_deviation` $p\ X$ = `sqrt` (`variance` $p\ X$)

[probabilityTheory.variance_alt]
$\vdash$ `variance` $p\ X$ =
`expectation` $p$ ($\lambda x$. ($X\ x$ − `expectation` $p\ X)^2$)

# Finite Second Moments

$\vdash$ finite_second_moments $p$ $X$ $\iff$
$\exists a$. second_moment $p$ $X$ $a$ $<$ $+\infty$

$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\Rightarrow$
(finite_second_moments $p$ $X$ $\iff$
second_moment $p$ $X$ $0$ $<$ $+\infty$)
$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\Rightarrow$
(finite_second_moments $p$ $X$ $\iff$
$\forall r$. second_moment $p$ $X$ (Normal $r$) $<$ $+\infty$)
$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\Rightarrow$
(finite_second_moments $p$ $X$ $\iff$ variance $p$ $X$ $<$ $+\infty$)
$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\Rightarrow$
(finite_second_moments $p$ $X$ $\iff$
integrable $p$ $(\lambda x.\ (X\ x)^2)$)
$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\Rightarrow$
(finite_second_moments $p$ $X$ $\iff$
$\forall c$. integrable $p$ $(\lambda x.\ (X\ x$ $-$ Normal $c)^2)$)

$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\wedge$
finite_second_moments $p$ $X$ $\Rightarrow$
expectation $p$ $X$ $\neq$ $+\infty$ $\wedge$ expectation $p$ $X$ $\neq$ $-\infty$
$\vdash$ prob_space $p$ $\wedge$ real_random_variable $X$ $p$ $\wedge$
finite_second_moments $p$ $X$ $\Rightarrow$
integrable $p$ $X$

# Basic Properties of Variance

$\vdash$ `prob_space` $p \Rightarrow 0 \leq$ `variance` $p$ $X$
$\vdash$ `prob_space` $p \Rightarrow$ `variance` $p$ $(\lambda x.$ `Normal` $c)$ = $0$
$\vdash$ `prob_space` $p \wedge$ `real_random_variable` $X$ $p \wedge$
`finite_second_moments` $p$ $X \Rightarrow$
`variance` $p$ $(\lambda x.$ `Normal` $c \times X$ $x)$ =
`Normal` $c^2 \times$ `variance` $p$ $X$
$\vdash$ `prob_space` $p \wedge$ `real_random_variable` $X$ $p \wedge$
`integrable` $p$ $X \wedge c \neq +\infty \wedge c \neq -\infty \Rightarrow$
`variance` $p$ $(\lambda x.$ $X$ $x + c)$ = `variance` $p$ $X$

$$D[X] = E[X^2] - E[X]^2$$

$\vdash$ `prob_space` $p \wedge$ `real_random_variable` $X$ $p \wedge$
`integrable` $p$ $(\lambda x.$ $(X$ $x)^2) \Rightarrow$
`variance` $p$ $X$ =
`expectation` $p$ $(\lambda x.$ $(X$ $x)^2) -$ (`expectation` $p$ $X)^2$
$\vdash$ `prob_space` $p \wedge$ `real_random_variable` $X$ $p \wedge$
`integrable` $p$ $(\lambda x.$ $(X$ $x)^2) \Rightarrow$
`variance` $p$ $X \leq$ `expectation` $p$ $(\lambda x.$ $(X$ $x)^2)$

# Uncorrelated r.v.'s and Covariance

$\vdash$ `uncorrelated` $p$ $X$ $Y$ $\iff$
`finite_second_moments` $p$ $X$ $\wedge$
`finite_second_moments` $p$ $Y$ $\wedge$
`expectation` $p$ $(\lambda s.\ X\ s \times Y\ s)$ =
`expectation` $p$ $X$ $\times$ `expectation` $p$ $Y$

$\vdash$ `covariance` $p$ $X$ $Y$ =
`expectation` $p$
$(\lambda x.$
$(X\ x\ -$ `expectation` $p$ $X) \times$
$(Y\ x\ -$ `expectation` $p$ $Y))$

$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X$ $p$ $\wedge$
`real_random_variable` $Y$ $p$ $\wedge$ `uncorrelated` $p$ $X$ $Y$ $\Rightarrow$
`expectation` $p$
$(\lambda s.$
$(X\ s\ -$ `expectation` $p$ $X) \times$
$(Y\ s\ -$ `expectation` $p$ $Y))$ =
$0$

$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X$ $p$ $\wedge$
`real_random_variable` $Y$ $p$ $\wedge$ `uncorrelated` $p$ $X$ $Y$ $\Rightarrow$
`covariance` $p$ $X$ $Y$ = $0$

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mathrm{E}[X])\,(Y-\mathrm{E}[Y])] = \mathrm{E}[XY]-\mathrm{E}[X]\,\mathrm{E}[Y]$$

# Sum of Variance of Uncorrelated r.v.'s

$$\mathrm{D}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{D}[X_i]$$

$$\begin{array}{l}
\vdash \texttt{prob_space}\ p \wedge \texttt{FINITE}\ J \wedge \\
\quad (\forall i.\ i \in J \Rightarrow \texttt{real_random_variable}\ (X\ i)\ p) \wedge \\
\quad \texttt{uncorrelated_vars}\ p\ X\ J \Rightarrow \\
\quad \texttt{variance}\ p\ (\lambda x.\ \sum\ (\lambda n.\ X\ n\ x)\ J)\ \texttt{=} \\
\quad \sum\ (\lambda n.\ \texttt{variance}\ p\ (X\ n))\ J
\end{array}$$

Proof (with only two r.v.'s): Let $\overline{X} = X - \mathrm{E}[X]$, $\overline{Y} = Y - \mathrm{E}[Y]$,

$$\mathrm{D}[X+Y] = \mathrm{E}[(\overline{X}+\overline{Y})^2] = \mathrm{E}[\overline{X}^2 + \overline{Y}^2 + 2\overline{X}\cdot\overline{Y}] = \mathrm{D}[X] + \mathrm{D}[Y] + 0$$

(NOTE: $\mathrm{E}[\overline{X}\cdot\overline{Y}] = \mathrm{E}[(X - \mathrm{E}[X])\,(Y - \mathrm{E}[Y])] = 0$ when $X$ and $Y$ are uncorrelated.

# Markov Inequality and Chebyshev's Inequality

## Markov Inequality (in Probability Space): $\mathrm{P}\{c \leqslant |X|\} \leqslant \frac{\mathrm{E}[|X|]}{c}$

$\vdash$ `prob_space` $p$ $\wedge$ `integrable` $p$ $X$ $\wedge$ $0 < c$ $\wedge$
  $a \in$ `events` $p$ $\Rightarrow$
  `prob` $p$ $(\{x \mid c \leq$ `abs` $(X\ x)\} \cap a) \leq$
  $c^{-1} \times$ `expectation` $p$ $(\lambda x.$ `abs` $(X\ x) \times \mathbb{1}\ a\ x)$

## Chebyshev's Inequality: $\mathrm{P}\{t \leqslant |\overline{X}|\} \leqslant \frac{\mathrm{E}[\overline{X}^2]}{t^2}$

$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X$ $p$ $\wedge$
  `finite_second_moments` $p$ $X$ $\wedge$ $0 < t \Rightarrow$
  `prob` $p$
    $(\{x \mid t \leq$ `abs` $(X\ x -$ `expectation` $p\ X)\} \cap$
     `p_space` $p) \leq t^{2\ -1} \times$ `variance` $p$ $X$

$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X$ $p$ $\wedge$
  `finite_second_moments` $p$ $X$ $\wedge$ $0 < t$ $\wedge$
  $a \in$ `events` $p$ $\Rightarrow$
  `prob` $p$ $(\{x \mid t \leq$ `abs` $(X\ x -$ `Normal` $c)\} \cap a) \leq$
  $t^{2\ -1} \times$
  `expectation` $p$ $(\lambda x.\ (X\ x -$ `Normal` $c)^2 \times \mathbb{1}\ a\ x)$

# Independence of Events

Two events $E_1$ and $E_2$ are *independent* if:

$$\mathrm{P}\{E_1 \cap E_2\} = \mathrm{P}\{E_1\}\,\mathrm{P}\{E_2\}$$

A (infinite) number of events $E_1$, $E_2$, ... are independent if any finite subset:

$$\mathrm{P}\left\{\bigcap_{i=1}^{n} E_i\right\} = \prod_{i=1}^{n} \mathrm{P}\{E_i\}$$

$\vdash$ $\texttt{indep}\ p\ a\ b \iff$
  $a \in \texttt{events}\ p \land b \in \texttt{events}\ p \land$
  $\texttt{prob}\ p\ (a \cap b) = \texttt{prob}\ p\ a \times \texttt{prob}\ p\ b$

$\vdash$ $\texttt{pairwise\_indep\_events}\ p\ E\ J \iff$
  $\forall i\ j.$
    $i \in J \land j \in J \land i \neq j \Rightarrow \texttt{indep}\ p\ (E\ i)\ (E\ j)$

$\vdash$ $\texttt{indep\_events}\ p\ E\ J \iff$
  $\forall N.\ N \subseteq J \land N \neq \emptyset \land \texttt{FINITE}\ N \Rightarrow$
    $\texttt{prob}\ p\ (\bigcap\ (\texttt{IMAGE}\ E\ N)) = \prod\ (\texttt{prob}\ p \circ E)\ N$

$\vdash$ $(\forall n.\ n \in J \Rightarrow E\ n \in \texttt{events}\ p) \land$
  $\texttt{indep\_events}\ p\ E\ J \Rightarrow$
  $\texttt{pairwise\_indep\_events}\ p\ E\ J$

# Independence of Sets of Events

Two sets of events $q$ and $r$ are *independent* if for any $E_1 \in q$, $E_2 \in r$,

$$\mathrm{P}\{E_1 \cap E_2\} = \mathrm{P}\{E_1\} \times \mathrm{P}\{E_2\}$$

$\vdash$ $\texttt{indep\_sets}\ p\ q\ r \iff$
$\quad \forall s\ t.\ s \in q \wedge t \in r \Rightarrow \texttt{indep}\ p\ s\ t$
$\vdash$ $\texttt{pairwise\_indep\_sets}\ p\ A\ J \iff$
$\quad \forall i\ j.$
$\quad\quad i \in J \wedge j \in J \wedge i \neq j \Rightarrow$
$\quad\quad \texttt{indep\_sets}\ p\ (A\ i)\ (A\ j)$
$\vdash$ $\texttt{indep\_sets}\ p\ A\ J \iff$
$\quad \forall N.\ N \subseteq J \wedge N \neq \emptyset \wedge \texttt{FINITE}\ N \Rightarrow$
$\quad\quad \forall E.\ E \in N \longrightarrow A \Rightarrow$
$\quad\quad\quad \texttt{prob}\ p\ (\bigcap\ (\texttt{IMAGE}\ E\ N)) =$
$\quad\quad\quad \prod\ (\texttt{prob}\ p \circ E)\ N$

$\vdash$ $(\forall n.\ n \in J \Rightarrow A\ n \subseteq \texttt{events}\ p) \wedge \texttt{indep\_sets}\ p\ A\ J \Rightarrow$
$\quad \texttt{pairwise\_indep\_sets}\ p\ A\ J$

# Independence of Random Variables (r.v.'s)

Two r.v.'s $X\colon \Omega \to \mathcal{A}_1$ and $Y\colon \Omega \to \mathcal{A}_2$ are *independent* if any two preimages of $X$ and $Y$ are independent events.

```
⊢ indep_vars p X Y s t ⟺
    ∀a b.
      a ∈ subsets s ∧ b ∈ subsets t ⇒
      indep p (PREIMAGE X a ∩ p_space p)
        (PREIMAGE Y b ∩ p_space p)

⊢ indep_vars p X A J ⟺
    ∀E N.
      N ⊆ J ∧ N ≠ ∅ ∧ FINITE N ∧
      E ∈ N ⟶ subsets ∘ A ⇒
      prob p
        (⋂
           (IMAGE
              (λn. PREIMAGE (X n) (E n) ∩ p_space p)
              N)) =
      ∏
        (prob p ∘
         (λn. PREIMAGE (X n) (E n) ∩ p_space p)) N
```

# Properties of Independent r.v.'s

Total independence implies pairwise independence:

$\vdash$ `prob_space` $p$ $\wedge$
`(`$\forall i.\ i \in J \Rightarrow$ `random_variable (`$X\ i$`)` $p$ `(`$A\ i$`))` $\wedge$
`(`$\forall i.\ i \in J \Rightarrow$ `sigma_algebra (`$A\ i$`))` $\wedge$
`indep_vars` $p\ X\ A\ J \Rightarrow$
`pairwise_indep_vars` $p\ X\ A\ J$

$\mathrm{E}[XY] = \mathrm{E}[X] \cdot \mathrm{E}[Y]$ if $X$ and $Y$ are independent integrable r.v.'s

$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X\ p$ $\wedge$
`real_random_variable` $Y\ p$ $\wedge$
`indep_vars` $p\ X\ Y$ `Borel Borel` $\wedge$ `integrable` $p\ X$ $\wedge$
`integrable` $p\ Y \Rightarrow$
`expectation` $p$ `(`$\lambda x.\ X\ x \times Y\ x$`) =`
`expectation` $p\ X \times$ `expectation` $p\ Y$
$\vdash$ `prob_space` $p$ $\wedge$ `real_random_variable` $X\ p$ $\wedge$
`real_random_variable` $Y\ p$ $\wedge$
`finite_second_moments` $p\ X$ $\wedge$
`finite_second_moments` $p\ Y$ $\wedge$
`indep_vars` $p\ X\ Y$ `Borel Borel` $\Rightarrow$
`uncorrelated` $p\ X\ Y$

# Convergence Concepts of Random Sequences

Consider $X$ and a (countable) sequence of r.v.'s $\{X_n\}$, taking (finite) real values:

- $\{X_n\}$ is said to *converge almost everywhere (a.e.)* (to $X$) if:

$$\exists N \in \mathcal{N}.\, \forall \omega \in \Omega \setminus N.\, \lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ finite}$$

- $\{X_n\}$ is said to *converge in probability (in pr.)* (to $X$) if:

$$\forall \epsilon > 0.\, \lim_{n\to\infty} \mathrm{P}\{|X_n - X| > \epsilon\} = 0$$

- $\{X_n\}$ is said to *converge in $L^p$* $(0 < p < \infty)$ to $X$ if $X, X_n \in L^p$ and

$$\lim_{n\to\infty} \mathrm{E}[|X_n - X|^p] = 0 \quad \text{or} \quad \lim_{n\to\infty} \int_\Omega |X_n - X|^p \,\mathrm{d}\mu = 0$$

# Convergence Concepts - Formal Definitions

$\vdash$ $(X \longrightarrow Y)$ (almost_everywhere $p$) $\iff$
AE $x$::$p$.
  $((\lambda n.$ real $(X\ n\ x)) \longrightarrow$ real $(Y\ x))$
    sequentially [converge_AE_def]

$\vdash$ $(X \longrightarrow Y)$ (in_probability $p$) $\iff$
$\forall e.\ 0 < e \land e \neq +\infty \Rightarrow$
  $((\lambda n.$
      real
        (prob $p$
          $\{x$ |
            $x \in$ p_space $p \land$
            $e <$ abs $(X\ n\ x - Y\ x)\}))$ $\longrightarrow 0)$
    sequentially [converge_PR_def]

$\vdash$ $(X \longrightarrow Y)$ (in_lebesgue $r$ $p$) $\iff$
$0 < r \land r \neq +\infty \land$
$(\forall n.$ expectation $p$ $(\lambda x.$ abs $(X\ n\ x)$ powr $r) \neq +\infty) \land$
expectation $p$ $(\lambda x.$ abs $(Y\ x)$ powr $r) \neq +\infty \land$
$((\lambda n.$
    real
      (expectation $p$
        $(\lambda x.$ abs $(X\ n\ x - Y\ x)$ powr $r)))$ $\longrightarrow 0)$
  sequentially [converge_LP_def]

# Relations between Convergence Concepts

- Convergence a.e. implies convergence in pr.:

$\vdash$ `prob_space` $p$ $\wedge$
($\forall n$. `real_random_variable` ($X$ $n$) $p$) $\wedge$
`real_random_variable` $Y$ $p$ $\wedge$
($X \longrightarrow Y$) (`almost_everywhere` $p$) $\Rightarrow$
($X \longrightarrow Y$) (`in_probability` $p$) [converge_AE_imp_PR]

$\vdash$ `prob_space` $p$ $\wedge$
($\forall n$. `real_random_variable` ($X$ $n$) $p$) $\wedge$
($X \longrightarrow (\lambda x.\ 0)$) (`almost_everywhere` $p$) $\Rightarrow$
($X \longrightarrow (\lambda x.\ 0)$) (`in_probability` $p$) [converge_AE_imp_PR']

- Convergence in $L^p$ implies convergence in pr.:

$\vdash$ `prob_space` $p$ $\wedge$
($\forall n$. `real_random_variable` ($X$ $n$) $p$) $\wedge$
($X \longrightarrow (\lambda x.\ 0)$) (`in_lebesgue` (&$k$) $p$) $\Rightarrow$
($X \longrightarrow (\lambda x.\ 0)$) (`in_probability` $p$) [converge_LP_imp_PR']

# Laws of Large Numbers (LLN)

Let $\{X_n\}$ be a random sequence and $\{S_n\}$ be the random sequence of partial sums

$$S_n = \sum_{i=1}^{n} X_n$$

The so-called "law of large numbers" says, under various conditions, the convergence (in a.e. or pr.):

$$\frac{S_n - \mathrm{E}(S_n)}{n} \longrightarrow 0$$

$\vdash$ $\texttt{LLN}\ p\ X\ \mathit{convergence\_mode} \iff$
  (**let**
    $Z\ n\ x = \sum (\lambda i.\ X\ i\ x)\ (\texttt{count1}\ n)$
  **in**
    $((\lambda n\ x.$
      $(Z\ n\ x - \texttt{expectation}\ p\ (Z\ n))\ /\ \texttt{\&SUC}\ n) \longrightarrow$
    $(\lambda x.\ 0))\ (\mathit{convergence\_mode}\ p))$   [large_numberTheory.LLN_def]

# The Weak Law of Large Numbers (WLLN)

For uncorrelated r.v.'s with a common bounded of variances:

$\vdash$ prob_space $p$ $\wedge$
  ($\forall n$. real_random_variable ($X$ $n$) $p$) $\wedge$
  ($\forall i$ $j$. $i \neq j \Rightarrow$ uncorrelated $p$ ($X$ $i$) ($X$ $j$)) $\wedge$
  ($\exists c$. $c \neq +\infty$ $\wedge$ $\forall n$. variance $p$ ($X$ $n$) $\leq c$) $\Rightarrow$
  LLN $p$ $X$ in_probability    [WLLN_uncorrelated]

Proof (through convergence in $L^2$): Let $M_n = E(S_n)$,

$$E\left[\left(\frac{S_n - M_n}{n} - 0\right)^2\right] = \frac{E[(S_n - M_n)^2]}{n^2} =$$

$$\frac{\mathrm{D}[S_n]}{n^2} = \frac{\sum_{i=1}^{n} \mathrm{D}[X_i]}{n^2} \leqslant \frac{nc}{n^2} = \frac{c}{n} \longrightarrow 0 \quad \text{as } n \to \infty$$

Thus $\frac{S_n - M_n}{n}$ converges to $0$ in $L^2$, thus also in pr.

# Other versions of Law of Large Numbers

```
⊢ prob_space p ∧
  (∀n. real_random_variable (X n) p) ∧
  (∀i j. i ≠ j ⇒ uncorrelated p (X i) (X j)) ∧
  (∃c. c ≠ +∞ ∧ ∀n. variance p (X n) ≤ c) ⇒
  LLN p X almost_everywhere                    [WLLN_uncorrelated]

⊢ identical_distribution p X E J ⟺
  ∀i j s.
    s ∈ subsets E ∧ i ∈ J ∧ j ∈ J ⇒
    distribution p (X i) s =
    distribution p (X j) s
⊢ prob_space p ∧
  (∀n. real_random_variable (X n) p) ∧
  pairwise_indep_vars p X (λn. Borel) 𝓤(:num) ∧
  identical_distribution p X Borel 𝓤(:num) ∧
  integrable p (X 0) ⇒
  LLN p X in_probability                                [WLLN_IID]
⊢ prob_space p ∧
  (∀n. real_random_variable (X n) p) ∧
  pairwise_indep_vars p X (λn. Borel) 𝓤(:num) ∧
  identical_distribution p X Borel 𝓤(:num) ∧
  integrable p (X 0) ⇒
  LLN p X almost_everywhere                             [SLLN_IID]
```