Some properties of probability theory in reverse mathematics

NingNing Peng

National University of Singapore
matpn@nus.edu.sg

Sep. 4, 2014
IMS-JSPS Joint Workshop in Mathematical Logic and Foundations of Mathematics

Outline

1. Reverse Mathematics on Measure Theory
   - The Borel-Cantelli lemma
   - Application
2. Convergences of random variables

Motivation and abstract

We want to find natural statements which are equivalent to WWKL. Since the WWKL is given in term of measure over Cantor space, it may be easir to get such statements in probability theory than others.

In this talk, we will deal with some basic properties in probability theory, and give some results in the context of reverse mathematics.

V. Reverse Mathematics on Measure Theory

The measure

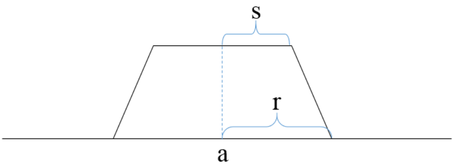
A complete separable metric space $\hat{A} = (A, d)$ is coded by the countable dense set $A \subset \mathbb{N}$ and the pseudo-metric d on A. A point of $\hat{A}$ is a strong Cauchy sequence $\langle a_n : n \in \mathbb{N} \rangle$ in the sense that $d(a_n, a_m) \leq 2^{-n}$ for any $n \leq m$.

A complete separable metric space $\hat{A}$ is compact if there exists an infinite sequence of finite sequences of $\hat{A}$, $\langle \langle x_{ij} : i \leq n_j \rangle : j \in \mathbb{N} \rangle$, such that for all $z \in \hat{A}$ and $j \in \mathbb{N}$ there exists $i \leq n_j$ such that $d(x_{ij}, z) < 2^{-j}$.

A basic open ball for this space $\hat{A}$ is coded by $\langle a, r \rangle$, where $a \in A$ and $r > 0$ is a rational. Open and closed subsets are coded by sequences of basic open balls.

A basic function on $\hat{A}$ is a code $p = \langle a, r, s \rangle$ where $a \in A$ and $r, s \in \mathbb{Q}$ which satisfy that $0 \leq s < r$. It is understood that $p = \langle a, r, s \rangle$ can be seen as a continuous function such that for any $x \in \hat{A}$,

$$p(x) = \begin{cases} 1 & \text{if } \ d(a,x) \leq s, \\ (r - d(a,x))/(r-s) & \text{if } \ s < d(a,x) < r, \\ 0 & \text{if } \ d(a,x) \geq r. \end{cases}$$



$C(\hat{A})$: the space of continuous real-valued functions

Let P be the set of linear combinations of basic functions with rational coefficients. Then P forms a vector space over the rational field.

The space of continuous real-valued functions $C(\hat{A})$ is defined as the complete separable Banach space $\hat{P} = (P, ||\cdot||_\infty)$. If $f = \langle p_n : n \in \mathbb{N} \rangle \in C(\hat{A})$, it is understood that $f(x) = \lim_n p_n(x)$ for any $x \in \hat{A}$.

Measure

A (probability) measure is a (code for a) positive linear functional μ on $C(\hat{A})$ such that $\mu(1) = 1$. For any open subset U of $\hat{A}$, the measure of U is defined to be

$$\mu(U) = \sup\{\mu(g) : g \prec U\},$$

where $g \prec U$ is used for the statement that $0 \leq g \leq 1$ and $g(x) = 0$ for any $x \notin U$. Similarly, for any closed subset C of X,

$$\mu(C) = \inf\{\mu(g) : C \prec g\}$$

In [3], Yu and Simpson give some basic properties.

Lemma (Yu and Simpson, [3])

The following can be proved in RCA_0,

- $\mu(\hat{A}) = 1$ *and* $\mu(\emptyset) = 0$.
- *If* U *is open and* $C = X \setminus U$, *then* $\mu(U) = 1 - \mu(C)$.
- *For any open sets* U *and* V, $U \subset V$ *implies that* $\mu(U) \leq \mu(V)$.
- *For any open sets* U *and* V, $\mu(U) + \mu(V) = \mu(U \cup V) + \mu(U \cap V)$.
- *For any open sets* U, $\mu(U) = \sup\{\mu(C) : C$ *is closed and* $C \subseteq U\}$.
- *For any open sets* C, $\mu(C) = \inf\{\mu(U) : U$ *is open and* $C \subseteq U\}$.

WWKL

Definition

We define weak weak König's lemma to be the following axiom:
if T is a subtree of $2^{<\mathbb{N}}$ with no infinite path, then

$$\lim_{n\to\infty} \frac{|\{\sigma \in T | lh(\sigma) = n\}|}{2^n} = 0.$$

$WWKL_0$ is a subsystem of Z_2 consisting of RCA_0 plus weak weak konig's lemma.

Recall some results

Theorem (countable additivity, Yu and Simpson, 1990, [3])

The following assertions are pairwise equivalent over RCA_0.

- WWKL.
- *(countable additivity) For any compact separable metric space* X *and any measure* μ *on* X, μ *is countably additive.*
- *For any covering of the closed unit interval* $[0, 1]$ *by a sequence of open intervals* (a_n, b_n), $n \in \mathbb{N}$, *we have* $\sum_{n=0}^{\infty} |a_n - b_n| \geq 1$.

Theorem (Yu, 1994, [2])

The following are pairwise equivalent over RCA_0.

1. *Weak weak König's lemma.*
2. *The monotone convergence theorem for measures on compact metric spaces.*
3. *The monotone convergence theorem for Lebesgue measure on* $[0,1]$.

Theorem (Yu, 1994, [2])

The following assertions are pairwise equivalent over RCA_0.

(i) ACA_0.

(ii) *Lebesgue domainated convergence theorem: If* $< f_n : n \in \mathbb{N} > \subset L^1(\mu)$ *is dominated convergent, then there is* $f \in L^1(\mu)$ *such that*

$$\lim_n \| f - f_n \|_1 = 0 \text{ and } \lim_n \int f_n d\mu = \int f d\mu.$$

Lebesgue dominated convergence theorem for the Lebesgue measure μ_L *on* $[0, 1]$.

The Borel-Cantelli lemma is a theorem about sequences of events, named after Emile Borel and Francesco Paolo Cantelli, who found it in the first decades of the 20th century.

Theorem

The following statement, called the first Borel-Cantelli lemma is equivalent to WWKL *over* RCA_0*: Let* O_n *be the sequences of open set. If* $\sum_{n=0} \mu(O_n) < \infty$*, then*

$$\mu(\bigcap_n \bigcup_{k>n} O_k) = 0.$$

A related result, sometimes called the second Borel-Cantelli lemma, is a partial converse of the first Borel-Cantelli lemma.
The sequences of $\langle O_n : n \in \mathbb{N} \rangle$ is independent if $\mu(\bigcap_{i<m} O_{n_i}) = \Pi_{i<m}\mu(O_{n_i})$ for all $\langle n_i : i < m \rangle$.

Theorem (RCA_0)

Let O_n be an independent sequence of open sets. If $\sum_{n=0} \mu(O_n) = \infty$, then

$$\mu(\bigcap_n \bigcup_{k>n} O_k) = 1.$$

Application of BC-lemma

Theorem

The following assertions are pairwise equivalent over RCA_0.

(i) ACA_0.

(ii) *Let* $\langle f_n : n \in \mathbb{N} \rangle$ *be a dominated Cauchy sequence of* $L_1(\hat{A}, \mu)$ *in probability. Then, there exists subsequences which converge to some* $f \in L_1(\hat{A}, \mu)$ *a.e.*

Proof

$(i \to ii)$ Give a sequence $\langle f_n : n \in \mathbb{N} \rangle$, where $f_n = \langle p_{n,l} : l \in \mathbb{N} \rangle \in L_1(\hat{A}, \mu)$. Assume that this sequence is dominated and cauchy in probability, that is,

$$\forall \varepsilon > 0, \lim_{n,m \to \infty} \mu(\{x \in \hat{A} : |f_n(x) - f_m(x)| > \varepsilon\}) = 0.$$

By ACA, we let $g(n)$ be the least m such that $\forall l \geq m$,

$$\mu(\{x \in \hat{A} : |p_{l,n+2}(x) - p_{m,n+2}(x)| > 2^{-n-2}\}) < 2^{-n}.$$

Then define $h : \mathbb{N} \to \mathbb{N}$ by $h(0) = g(0)$ and $h(n+1) = g(m)$ where m is the least number s.t $g(m) > h(n)$.

Continue of Proof

Let $E_k = \{x \in \Omega : |P_{h(k+1),k+2}(x) - P_{h(k),k+2}(x)| > 2^{-k-2}\}$. Since $\mu(E_k) < 2^{-k}$, by the first Borel-Cantelli lemma, $\mu(\bigcap_n \bigcup_{k \geq n} E_k) = 0$. So, $\langle f_{h(k)} : k \in \mathbb{N} \rangle$ is point-wise convergent a.e.
By Lebesgue dominated convergence theorem, $\exists f \in L_1(\hat{A}, \mu)$ such that $f_{h(k)} \to f$ a.e., that is, in $|| \ ||_1$.

$(ii \to i)$ As the proof of Theorem III 2.2 in Simpson's book.

II. Convergence in probability

For a compact metric space Ω and a probability measure P on Ω, the pair (Ω, P) is said to be a *sample space*.
When we say X is a *random variable*, then $X \in C(\Omega)$.
For any reals $a < b$ and any random variable X, we define

$$P(\{a < X < b\}) = P(\{\omega \in \Omega : a < X(\omega) < b\}).$$

Let $m = E(X) = P(X)$. Since $(X - m)^2 \in C(\Omega)$, we can define $V(X) := E((X - m)^2)$.

Fix a sample space.

Definition (RCA_0)

Let $\langle X_n : n \in \mathbb{N} \rangle$ be a sequence of random variables and X a random variable.

- X_n converges to X almost surely if for all $\epsilon > 0$, $P(\cap_n \cup_{k \geq n} \{|X_n - X| > \epsilon\}) = 0$, that is, $\lim_{n\to\infty} P(\cup_{k\geq n}\{|X_n - X| > \epsilon\}) = 0$.
- X_n converges to X in probability if for all $\epsilon > 0$, $\lim_{n\to\infty} P(\{|X_n - X| > \epsilon\}) = 0$.
- X_n converges to X in quadratic mean if $\lim_{n\to\infty} \| X_n - X \|_2 = 0$.
- X_n converges to X in L^1 if $\lim_{n\to\infty} \| X_n - X \|_1 = 0$.
- F_X is a (continuous) distribution function of X if $P(X \leq t) = F_X(t)$.
- X_n converges to X in distribution if $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ for all t.

Proposition (RCA_0)

Let $\langle X_n : n \in \mathbb{N}\rangle$ be a sequence or randome variables and X a random variable.

- *$X_n \to X$ a.s. implies $X_n \to X$ in probability.*
- *$X_n \to X$ in L^2 implies $X_n \to X$ in L^1.*
- *$X_n \to X$ in L^1 implies $X_n \to X$ in probability.*
- *If X_n's and X have distribution functions, then, $X_n \to X$ in probability implies $X_n \to X$ in distribution.*
- *If X_n's and X have distribution functions, $X_n \to c$ in distribution implies $X_n \to c$ in probability, where c is constant.*

Theorem

The following assertions are pairwise equivalent over RCA_0.

(i) WWKL_0.

(ii) $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$ *for every* $\epsilon > 0$ *implies* $X_n \to X$ *a.s.*

Proof.

We prove (ii) to (i). Let $\langle (a_i, b_i) : i \in \mathbb{N} \rangle$ be a sequence of open intervals such that $\sum_{i=1}^{\infty} P((a_i, b_i)) < 1$.
Define basic functions $\langle X_i : i \in \mathbb{N} \rangle$ by

$$X_i = \langle (a_i + b_i)/2, (b_i - a_i)/2, (b_i - a_i)/2 + 2^{-i} \rangle$$

By (ii), we get $X_n \to 0$ a.s. Then for any $0 < \varepsilon < 1$,
$\lim_{n\to\infty} P(\cup_{i>n}(a_i, b_i)) \leq \lim_{n\to\infty} P(\cup_{k\geq n}\{|X_n| > \epsilon\}) = 0$. This gives an instance of Borel-Cantelli Lemma to prove WWKL. □

Lemma (WKL_0)

Let X_n, X be with distribution functions. Assume that $|X_n|$'s and $|X|$ are bounded by $M > 0$. Then, $X_n \to X$ in distribution if and only if $E[f(X_n)] \to E[f(X)]$ for any continuous function $f : [-M, M] \to \mathbb{R}$.

Lemma ($WWKL_0$)

Convergence in probability implies there exists a sub-sequence which almost surely converges.

Proof.

Without loss of generality, we may assume that for all $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n| \geq \varepsilon) = 0$.

Since $P(|X_n| \geq 2^{-k}) < 2^{-k}$ is Σ^0_1, there exists a function $f : \mathbb{N} \to \mathbb{N}$ such that $f(k+1) > f(k)$ and $P(|X_{f(k)}| \geq 2^{-k}\}) < 2^{-k}$.

Define an open set $E_k = \{\omega : |X_{f(k)}(\omega)| > 2^{-k}\}$. Since $P(E_k) \leq 2^{-k}$, then $\sum_{k=0}^{\infty} P(E_k) < \infty$. By the first Borel-Cantelli lemma, $\lim_{n\to\infty} P(\bigcup_{k\geq n} E_k) = 0$.

So $\lim_{n\to\infty} P(\cup_{k\geq n}\{|X_{f(k)}| > \epsilon\}) \leq \lim_{n\to\infty} P(\bigcup_{k\geq n} E_k) = 0$ □

References

- Stephen G. Simpson: *Subsystems of Z_2 and reverse mathematics. Appendix to: G. TAKEUTI,*
  Proof Theory, 2nd ed. North-Holland Publ. Comp., Amsterdam 1986, pp. 434 - 448.
- Xiaokang Yu: *Lebesgue Convergence Theorems and Reverse Mathematics,*
  Math. Log. Quart. 40 (1994) 1-13.
- Xiaokang Yu and Stephen G. Simpson: *Measure theory and weak König's lemma*
  Archive Math. Logic, 30 (1990), 171-180.
- Brown, Douglas K., Mariagnese Giusto, and Stephen G. Simpson: *Vitali's theorem and WWKL*
  Archive for Mathematical Logic 41.2 (2002): 191-206.

Thank you very much!