Information-Theoretic Proof of the Hewitt-Savage zero-one law

Neil O' Connell
Basic Research Institute in Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-2000-18
12th July, 2000*

information theory, Hewitt-Savage zero-one law

In this short note we give an elementary information-theoretic proof of the Hewitt-Savage zero-one law.

* Internal Accession Date Only
Approved for External Publication

Information-theoretic proof of the Hewitt-Savage zero-one law

Neil O'Connell [1], *June 2000*

Abstract

In this short note we give an elementary information-theoretic proof of the Hewitt-Savage zero-one law.

Keywords: Information theory, Hewitt-Savage zero-one law.

We will use the notation of [2]. The following elementary inequality is at the heart of this discussion.

Lemma 1 *Let $X_1, X_2, \ldots$ be a sequence of independent, discrete random variables, and Z another discrete random variable on the same probability space. Then*

$$H(Z) \geq \sum_{i=1}^{\infty} I(X_i; Z).$$

Proof.

$$\begin{aligned} H(Z) &\geq H(Z) - H(Z|X_1, \ldots, X_n) = I(Z; X_1, \ldots, X_n) \\ &= \sum_{i=1}^{n} I(X_i; Z|X_{i-1}, \ldots, X_1) \\ &= \sum_{i=1}^{n} H(X_i) - H(X_i|Z, X_{i-1}, \ldots, X_1) \\ &\geq \sum_{i=1}^{n} H(X_i) - H(X_i|Z) = \sum_{i=1}^{n} I(X_i; Z). \end{aligned}$$

Now let $n \to \infty$. □

Now let $Y_1, Y_2, \ldots$ be a sequence of *i.i.d.* random variables taking values in a finite set Σ. The exchangeable σ-field $\mathcal{E}$ is the set of events which are invariant under finite permutations of the indices in the sequence Y.

[1] Postal address: BRIMS, HP Labs, Bristol BS34 8QZ, UK

Theorem 2 [Hewitt-Savage] *If $A \in \mathcal{E}$, then $P(A) \in \{0, 1\}$.*

Proof. Let $A \in \mathcal{E}$. By Lemma 1, for each $k = 1, 2, \ldots$,

$$H(1_A) \geq \sum_{i=0}^{\infty} I(1_A; Y_{ik+1}, \ldots, Y_{(i+1)k}). \tag{1}$$

Now $A \in \mathcal{E}$, so $I(1_A; Y_{ik+1}, \ldots, Y_{(i+1)k})$ does not depend on i. Since $H(1_A)$ must be finite, it follows that

$$I(1_A; Y_{ik+1}, \ldots, Y_{(i+1)k}) = 0,$$

for all i and k. In particular,

$$I(1_A; Y_1, \ldots, Y_n) = 0 \tag{2}$$

for all n. In other words, the event A is independent of $\sigma(Y_1, \ldots, Y_n)$, for all n, and hence independent of $\sigma(Y_1, Y_2, \ldots)$. But $A \in \sigma(Y_1, Y_2, \ldots)$, so A is independent of itself, and the result follows. □

This approach can be easily extended to the case where Σ is an arbitrary measurable space equipped with a countably generated σ-field. This is the case if, for example, Σ is a standard Borel space, or a Borel subset of a Polish space, or simply $\mathbb{R}^d$.

In the proof of Theorem 2 given in [1], the key observation (2) is deduced from the Kolmogorov zero-one law, which states that the tail σ-field

$$\mathcal{T} = \lim_{n \to \infty} \sigma(Y_n, Y_{n+1}, \ldots)$$

is trivial.

Finally we remark that Lemma 1 can also be regarded as an information-theoretic formulation of the following interpretation of the law of averages [4]:

> A quantity which depends, in a relatively uniform manner, on many independent random variables, is essentially constant.

To see this, observe that, in the proof of Theorem 2 given above, it would be possible to reach the same conclusion if $I(1_A; Y_{ik+1}, \ldots, Y_{(i+1)k})$ does not depend *too much* on i, in any sense which would force the sum in (1) to be infinite unless all of the terms are identicallly zero.

In [4] there is a discussion on concentration inequalities as a natural formulation of the law of averages. See also [3] for a similar discussion on the contraction principle of large deviation theory.

References

[1] P. Billingsley. *Probability and Measure.* Wiley, New York, 1979.

[2] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley (Series in Telecommunications) 1991.

[3] Neil O'Connell and Adam Shwartz. Tunneling from random walks to Markov chains. BRIMS Technical Report HPL-BRIMS-9809.

[4] M. Talagrand. A new look at independence. *Ann. Probab.* **24** 1-24, 1996.