## Randomness Deficiency Overlap

Samuel Epstein\*

July 5, 2023

## Abstract

In this paper we prove a lower bound on the computable measure of sets with high randomness deficiency with respect to two computable measures.

In this paper, we show a succinct proof to the randomness deficiency overlap theorem. The proof is a straightforward modification to Theorems 4, 5, and 6 in [Eps22]. In [Eps23], an extended version of the results of this paper can be found. It includes extensions for uncomputable  $\lambda$  as well as a statement using universal uniform tests and computable metric spaces. It also includes a result (with tight bounds) using the traditional definition of randomness deficiency. The paper also proves that synchronized oscillation of algorithmic thermodynamic entropies with respect to different measures must occur.

Let  $\mu = \mu_1, \mu_2, \ldots$  be a computable sequence of measures over infinite sequences. A conditionally bounded  $\mu$ -test is a lower semi-continuous function  $t : \{0,1\}^{\infty} \times \mathbb{N} \to \mathbb{R}_{\geq 0} \cup \infty$  such that for all  $n \in \mathbb{N}$  and positive real number r, we have  $\mu_n(\{\alpha : t(x|n) \geq r\}) \leq 1/r$ . If  $\mu_1, \mu_2, \ldots$  is uniformly computable, then there exists a lower-semicomputable  $\mu$ -test t that is "maximal" (i.e. for which  $t' \leq O(t)$  for every other test t'). We fix such a t and let  $\overline{\mathbf{D}}_n(\alpha|\mu) = \log t(\alpha|n)$ .

**Theorem.**Let  $\lambda = \lambda_1, \lambda_2, \ldots, \mu = \mu_1, \mu_2, \ldots,$  and  $\nu = \nu_1, \nu_2, \ldots$  be three uniformly computable sequences of measures over infinite sequences. Each  $\lambda_n$  is non-atomic. There is a constant  $c \in \mathbb{N}$ , where for all  $n \in \mathbb{N}$ ,  $\lambda_n \left\{ \alpha : \overline{\mathbf{D}}_n(\alpha | \mu) > n - c \text{ and } \overline{\mathbf{D}}_n(\alpha | \nu) > n - c \right\} > 2^{-n-1}$ .

## 1 Results

A sampling method A is a probabilistic function that maps an integer N with probability 1 to a set containing N different strings.

**Lemma 1** Let P and Q be two probability measures on strings and let A be a sampling method. For all integers N, there exists a finite set  $S \subset \{0,1\}^*$  such that  $P(S) \leq 32/N$ ,  $Q(S) \leq 32/N$ , and with probability strictly more than 0.99: A(N) intersects S.

**Proof.** We show that some possibly infinite set S satisfies the conditions, and thus, some finite subset also satisfies the conditions due to the strict inequality. We use the probabilistic method: we select each string to be in S with probability 8/N and show that the three conditions are satisfied with positive probability. The expected value of P(S) and Q(S) is 8/N. By the Markov inequality, the probability that P(S) > 32/N is at most 1/4 and the probability that Q(S) > 32/N is at most 1/4. For any set D containing N strings, the probability that S is disjoint from S is

$$(1 - 8/N)^N < e^{-8}.$$

<sup>\*</sup>JP Theory Group. samepst@jptheorygroup.org

Let Q be the measure over N-element sets of strings generated by the sampling algorithm A(N). The left-hand side above is equal to the expected value of

$$Q({D:D \text{ is disjoint from } S}).$$

Again by the Markov inequality, with probability greater than 3/4, this measure is less than  $4e^{-8} < 0.01$ . By the union bound, the probability that at least one of the conditions is violated is less than 1/4 + 1/4 + 1/4. Thus, with positive probability a required set is generated, and thus such a set exists.

Let  $P=P_1,P_2,...$  be a sequence of measures over strings. For example, one may choose  $P_1=P_2...$  or choose  $P_n$  to be the uniform measure over n-bit strings. A conditional probability bounded P-test is a function  $t:\{0,1\}^*\times\mathbb{N}\to\mathbb{R}_{\geq 0}$  such that for all  $n\in\mathbb{N}$  and positive real number r, we have  $P_n(\{x:t(x|n)\geq r\})\leq 1/r$ . If  $P_1,P_2,...$  is uniformly computable, then there exists a lower-semicomputable such P-test t that is "maximal" (i.e., for which  $t'\leq O(t)$  for every other such test t'). We fix such a t, and let  $\overline{\mathbf{d}}_n(x|P)=\log t(x|n)$ .

**Theorem 1** Let  $P = P_1, P_2...$  and  $Q = Q_1, Q_2...$  be a two uniformly computable sequence of measures on strings and let A be a sampling method. There exists  $c \in \mathbb{N}$  such that for all n:

$$\Pr\left(\max_{a\in A(2^n)}\min\{\overline{\mathbf{d}}_n(a|P),\overline{\mathbf{d}}_n(a|Q)\}>n-c\right)\geq 0.99.$$

**Proof.** We now fix a search procedure that on input N finds a set  $S_N$  that satisfies the conditions of Lemma 1. Let t'(a|n) be the maximal value of  $2^n/64$  such that  $a \in S_{2^n}$ . By construction, t' is a computable probability bounded test for both P and Q, because  $P_n(\{x:t'(x|n)=2^\ell\}) \leq 2^{-\ell-1}$ , and thus  $P_n(t'(x|n) \geq 2^\ell) \leq 2^{-\ell-1} + 2^{-\ell-2} + \dots$  and similarly for Q. With probability 0.99, the set  $A(2^n)$  intersects  $S_{2^n}$ . For any number a in the intersection, we have  $t'(x|n) \geq 2^{n-6}$ , thus by the optimality of t and definition of  $\overline{\mathbf{d}}$ , we have  $\overline{\mathbf{d}}_n(a|P) > n - O(1)$  and  $\overline{\mathbf{d}}_n(a|Q) > n - O(1)$ .

An incomplete sampling method A takes in a natural number N and outputs, with probability f(N), a set of N numbers. Otherwise A outputs  $\bot$ . f is computable.

**Corollary 1** Let  $P = P_1, P_2...$  and  $Q = Q_1, Q_2...$  be two uniformly computable sequences of measures on strings and let A be an incomplete sampling method. There exists  $c \in \mathbb{N}$  such that for all n:

$$\Pr_{D=A(n)}\left(D\neq \perp \ and \ \max_{a\in D}\min\{\overline{\mathbf{d}}_n(a|P),\overline{\mathbf{d}}_n(a|Q)\} \leq n-c\right) < 0.01.$$

A continuous sampling method C is a probabilistic function that maps, with probability 1, an integer N to an infinite encoding of N different sequences.

**Theorem 2** Let  $\mu = \mu_1, \mu_2, \ldots$  and  $\nu = \nu_1, \nu_2, \ldots$  be two uniformly computable sequences of measures over infinite sequences. Let C be a continuous sampling method. There exists  $c \in \mathbb{N}$  where for all n:

$$\Pr\left(\max_{\alpha \in C(2^n)} \min\{\overline{\mathbf{D}}_n(\alpha|\mu), \overline{\mathbf{D}}_n(\alpha|\nu)\} > n - c\right) \ge 0.98.$$

**Proof.** For  $D \subseteq \{0,1\}^{\infty}$ ,  $D_m = \{\omega[0..m] : \omega \in D\}$ . Let  $g(n) = \arg\min_{m} \Pr_{D=C(n)}(|D_m| < n) < 0.01$  be the smallest number m such that the initial m-segment of C(n) are sets of n strings with probability > 0.99. g is computable, because C outputs a set of distinct infinite sequences with probability 1. For probability  $\psi$  over  $\{0,1\}^{\infty}$ , let  $\psi^m(x) = [|x| = m]\psi(\{\omega : x \sqsubset \omega\})$ . Let  $\mu^g = \mu_1^{g(1)}, \mu_2^{g(2)}, \ldots$  and  $\nu^g = \nu_1^{g(1)}, \nu_2^{g(2)}, \ldots$  be two uniformly computable sequences of discrete probability measures and let A be a discrete incomplete sampling method, where for random seed  $\omega \in \{0,1\}^{\infty}$ ,  $A(n,\omega) = C(n,\omega)_{g(n)}$  if  $|C(n,\omega)_{g(n)}| = n$ ; otherwise  $A(n,\omega) = \bot$ . So  $\Pr[A(n) = \bot] < 0.01$ . There exists a constant  $c \in \mathbb{N}$  such that,

$$\Pr\left(\max_{\alpha \in C(2^{n})} \min\{\overline{\mathbf{D}}_{n}(\alpha|\mu), \overline{\mathbf{D}}_{n}(\alpha|\nu)\} \leq n - c\right)$$

$$\leq \Pr_{Z=C(2^{n})} \left( (|Z_{g(n)}| < 2^{n}) \text{ or } (|Z_{g(n)}| = 2^{n} \text{ and } \max_{\alpha \in Z} \min\{\overline{\mathbf{D}}_{n}(\alpha|\mu), \overline{\mathbf{D}}_{n}(\alpha|\nu)\} \leq n - c\right)$$

$$\leq \Pr_{D=A(2^{n})} \left( D = \bot \text{ or } (D \neq \bot \text{ and } \max_{x \in D} \min\{\overline{\mathbf{d}}_{n}(x|\mu^{g}), \overline{\mathbf{d}}_{n}(x|\nu^{g})\} \leq n - c\right)$$

$$< 0.01 + 0.01$$

$$< 0.02,$$

$$(1)$$

where Equation 1 is due to Corollary 1.

**Theorem 3** Let  $\lambda = \lambda_1, \lambda_2, \ldots, \mu = \mu_1, \mu_2, \ldots$ , and  $\nu = \nu_1, \nu_2, \ldots$  be three uniformly computable sequences of measures over infinite sequences. Each  $\lambda_n$  is non-atomic. There is a constant  $c \in \mathbb{N}$ , dependent on  $\mu$ ,  $\nu$  and  $\lambda$ , where for all  $n \in \mathbb{N}$ ,  $\lambda_n \{\alpha : \overline{\mathbf{D}}_n(\alpha|\mu) > n - c \text{ and } \overline{\mathbf{D}}_n(\alpha|\nu) > n - c\} > 2^{-n-1}$ .

**Proof.** We define the continuous sampling method C, where on input n, randomly samples n elements from  $\lambda_n$ . Let  $d_n = \lambda_n\{\alpha : \min\{\overline{\mathbf{D}}_n(\alpha|\mu), \overline{\mathbf{D}}_n(\alpha|\nu)\} > n-c\}$ , where c is the constant in Theorem 2. By that theorem,

$$\Pr\left(\max_{\alpha \in C(2^n)} \min\{\overline{\mathbf{D}}_n(\alpha|\mu), \overline{\mathbf{D}}_n(\alpha|\nu)\} > n - c\right) > 0.98$$

$$1 - (1 - d_n)^{2^n} > 0.98$$

$$1 - 2^n d_n < 0.02$$

$$d_n > (0.98)2^{-n}$$

$$\lambda_n\{\alpha : \min\{\overline{\mathbf{D}}_n(\alpha|\mu), \overline{\mathbf{D}}_n(\alpha|\nu)\} > n - c\} > 2^{-n-1}.$$

## References

[Eps22] S. Epstein. The outlier theorem revisited. CoRR, abs/2203.08733, 2022.

[Eps23] S. Epstein. Randomness Deficiency Overlap (Extended Version), 2023. http://www.jptheorygroup.org/doc/DeficiencyOverlap.pdf.