# Marginis

Bjørn Kjos-Hanssen

December 12, 2024

# Chapter 1

# Kjos-Hanssen (2010)

Excerpts from the paper *The probability distribution as a computational resource for randomness testing.*

## 1.1 Introduction

The fundamental idea of statistics is that by repeated experiment we can learn the underlying distribution of the phenomenon under investigation. In this paper we partially quantify the amount of randomness required to carry out this idea. We first show that ordinary Martin-Löf randomness with respect to the distribution is sufficient. Somewhat surprisingly, however, the picture is more complicated when we consider a weaker form of randomness where the tests are *effective*, rather than merely *effective relative to the distribution*. We show that such *Hippocratic* randomness actually coincides with ordinary randomness in that the same outcomes are random for each notion, but the corresponding test concepts do *not* coincide: while there is a universal test for ordinary ML-randomness, there is none for Hippocratic ML-randomness.

For concreteness we will focus on the classical Bernoulli experiment, although as the statistical tools we need are limited to Chebyshev's inequality and the strong law of large numbers, our result works also in the general situation of repeated experiments in statistics, where an arbitrary sequence of independent and identically distributed random variables is studied.

When using randomness as a computational resource, the most convenient underlying probability distribution may be that of a fair coin. In many cases, fairness of the proverbial coin may be only approximate. Imagine that an available resource generates randomness with respect to a distribution for which the probability of heads is $p \neq 1/2$. It is natural to assume that $p$ is not a computable number if the coin flips are generated with contributions from a physical process such as the flipping of an actual coin. The non-computability of $p$ matters strongly if an infinite sequence of coin flips is to be performed. In that case, the gold standard of algorithmic randomness is *Martin-Löf randomness*, which essentially guarantees that no algorithm (using arbitrary resources of time and space) can detect any regularities in the sequence. If $p$ is non-computable, it is possible that $p$ may itself be a valuable resource, and so the question arises whether a "truly random" sequence should look random even to an adversary equipped with the distribution as a resource. In this article we will show that the question is to some extent moot, as these types of randomness coincide. On the other hand, while there is a universal test for randomness in one case, in the other there is not. This article can be seen as a follow-up to Martin-Löf's paper where he introduced his notion of algorithmic randomness and proved results for Bernoulli measures [10].

It might seem that when testing for randomness, it is essential to have access to the distribution we are testing randomness for. On the other hand, perhaps if the results of the experiment are truly random we should be able to use them to discover the distribution for ourselves, and then once we know the distribution, test the results for randomness. However, if the original results are not really random, we may "discover" the wrong distribution. We show that there are tests that can be effectively applied, such that if the results are random then the distribution can be discovered, and the results will then turn out to be random even to someone who knows the distribution. While these tests can individually be effectively applied, they cannot be effectively enumerated as a family. On the other hand, there is a single such test (due to Martin-Löf) that will reveal whether the results are random for *some* (Bernoulli) distribution, and another (introduced in this paper) that if so will reveal that distribution.

In other words, one can effectively determine whether randomness for some distribution obtains, and if so determine that distribution. There is no need to know the distribution ahead of time to test for randomness with respect to an unknown distribution. If we suspect that a sequence is random with respect to a measure given by the value of a parameter (in an effective family of measures), there is no need to know the value of that parameter, as we can first use Martin-Löf's idea to test for randomness with respect to *some* value of the parameter, and then use the fundamental idea of statistics to *find* that parameter. Further effective tests can be applied to compare that parameter $q$ with rational numbers near our target parameter $p$, leading to the conclusion that if all *effective* tests for randomness with respect to parameter $p$ are passed, then all tests *having access to $p$ as a resource* will also be passed. But we need the distribution to know *which* effective tests to apply. Thus we show that randomness testing with respect to a target distribution $p$ can be done by two agents each having limited knowledge: agent 1 has access to the distribution $p$, and agent 2 has access to the data $X$. Agent 1 tells agent 2 which tests to apply to $X$.

The more specific point is that the information about the distribution $p$ required for randomness testing can be encoded in a set of effective randomness tests; and the encoding is intrinsic in the sense that the ordering of the tests does not matter, and further tests may be added: passing any collection of tests that include these is enough to guarantee randomness. From a syntactic point of view, whereas randomness with respect to $p$ is naturally a $\Sigma^0_2(p)$ class, our results show that it is actually an intersection of $\Sigma^0_2$ classes.

**Definition 1.** The Bernoulli measure $\mu_p$ is defined by the stipulation that for each $n \in \omega = \{0, 1, 2, ...\}$,

$$\mu_p(\{X : X(n) = 1\}) = p \text{ and}$$
$$\mu_p(\{X : X(n) = 0\}) = 1 - p$$

and $X(0), X(1), X(2), ...$ are mutually independent random variables.

If $X$ is a $\{0, 1\}$-valued random variable such that $\mathbb{P}(X = 1) = p$ then $X$ is called a Bernoulli($p$) random variable.

**Definition 2.** A $\mu_p$-ML-randomness test is a sequence $\{U^p_n\}_n$ that is uniformly $\Sigma^0_1(p)$ with $\mu(U^p_n) \leq 2^{-n}$, where $2^{-n}$ may be replaced by any computable function that goes to zero effectively.

A $\mu_p$-ML-randomness test is *Hippocratic* if there is a $\Sigma^0_1$ class $S \subseteq 2^\omega \times \omega$ such that $S = \{(X, n) : X \in U^p_n\}$. Thus, $U_n = U^p_n$ does not depend on $p$ and is uniformly $\Sigma^0_1$. If $X$ passes all $\mu_p$-randomness tests then $X$ is *$\mu_p$-random*. If $X$ passes all Hippocratic tests then $X$ is *Hippocrates $\mu$-random*.

To explain the terminology: like the ancient medic Hippocrates we are not consulting the oracle of Delphi ($p$) but rather looking for "natural causes". This level of randomness recently arose in the study of randomness extraction from subsets of random sets [8].

We will often write "$\mu_p$-random" instead of "$\mu_p$-ML-random", as we work in the Martin-Löf mode of randomness throughout, except when discussing a conjecture at the end of this paper.

## 1.2 Chebyshev's inequality

We develop this basic inequality from scratch here, in order to emphasize how generally it holds. For an event $A$ in a probability space, we let $\mathbf{1}_A$, the indicator function of $A$, equal 1 if $A$ occurs, and 0 otherwise. The *expectation* of a discrete random variable $X$ is

$$\mathbb{E}(X) = \sum_x x \cdot \mathbb{P}(X = x).$$

where $\mathbb{P}$ denotes probability and the sum is over all outcomes in the sample space. Thus $\mathbb{E}(X)$ is the average value of $X$ over repeated experiments. It is immediate that

$$\mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A).$$

Next we observe that the random variable that is equal to $a$ when a nonnegative random variable $X$ satisfies $X \geq a$ and 0 otherwise, is always dominated by $X$. That is,

$$a \cdot \mathbf{1}_{\{X \geq a\}} \leq X.$$

Therefore, taking expectations of both sides,

$$a \cdot \mathbb{P}\{X \geq a\} \leq \mathbb{E}(X).$$

In particular, for any random variable $X$ with $\mathbb{E}(X) = \mu \in \mathbb{R}$ we have

$$a^2 \cdot \mathbb{P}\{(X - \mu)^2 \geq a^2\} \leq \mathbb{E}((X - \mu)^2) =: \sigma^2$$

so

$$\mathbb{P}\{|X - \mu| \geq |a|\} \leq \sigma^2/a^2$$

If we let $k \in \omega$ and replace $a$ by $k\sigma$, then

$$\mathbb{P}\{|X - \mu| \geq k\sigma\} \leq \sigma^2/(k\sigma)^2 = 1/k^2.$$

This is Chebyshev's inequality, which in words says that the probability that we exceed the mean $\mu$ by $k$ many standard deviations $\sigma$ is rather small.

## 1.3 Results for ordinary randomness

We first prove a version of the phenomenon that for samples of sufficiently fast growing size, the sample averages almost surely converge quickly to the mean.

**Proposition 3.** *Consider a sequence $Y = \{Y_n\}_{n\in\omega}$ of independent Bernoulli(p) random variables, with the sample average*

$$\overline{Y}_n := \frac{1}{n}\sum_{i=0}^{n-1} Y_i.$$

*Let $N(b) = 2^{3b-1}$ and let*

$$U_d = \bigcup_{b\geq d}\{Y : |\overline{Y}_{N(b)} - p| \geq 2^{-b}\}.$$

*Then $U_d$ is uniformly $\Sigma^0_1(p)$, and $\mu_p(U_d) \leq 2^{-d}$, i.e., $\{U_d\}_{d\in\omega}$ is a $\mu_p$-ML-test.*

The idea of the proof is to use Chebyshev's inequality and the fact that the variance of a Bernoulli($p$) random variable is bounded (in fact, bounded by 1/4).

*Proof.* The fact that $U_d$ is $\Sigma^0_1(p)$ is immediate, so we prove the bound on its $\mu_p$-measure. We have

$$\mathbb{E}(\overline{Y}_n) = p \text{ and } \sigma^2(\overline{Y}_n) = \sigma^2/n$$

where $\sigma^2 = p(1-p) \leq 1/4$ is the variance of $Y_0$ and $\sigma^2(\overline{Y}_n)$ denotes the variance of $\overline{Y}_n$. Thus $\sigma \leq 1/2$, and

$$\mathbb{P}\left\{|\overline{Y}_n - p| \geq k \cdot \sigma(\overline{Y}_n)\right\} \leq 1/k^2,$$

so

$$\mathbb{P}\left\{|\overline{Y}_n - p| \geq \frac{k}{2\sqrt{n}}\right\} \leq \mathbb{P}\left\{|\overline{Y}_n - p| \geq \frac{k \cdot \sigma}{\sqrt{n}}\right\} \leq 1/k^2 =: 2^{-(b+1)}.$$

Now, we claim that $2^{-b} \geq \frac{k}{2\sqrt{n}}$ by taking $n$ large enough as a function of $b$:

$$n \geq k^2 4^{b-1} = 2^{b+1}4^{b-1} = 2^{3b-1}.$$

Thus, if $n \geq N(b) := 2^{3b-1}$,

$$\mathbb{P}\left\{|\overline{Y}_n - p| \geq 2^{-b}\right\} \leq 2^{-(b+1)}$$

so

$$\mathbb{P}\left\{\exists b \geq d \ |\overline{Y}_{N(b)} - p| \geq 2^{-b}\right\} \leq \sum_{b \geq d} 2^{-(b+1)} = 2^{-d}.$$

□

The following result in a sense encapsulates the essence of statistics.

**Theorem 4.** *If $Y$ is $\mu_p$-ML-random then $Y$ Turing computes $p$.*

*Proof.* We may assume $p$ is not computable, else there is nothing to prove; in particular we may assume $p$ is not a dyadic rational.

Let $\{U_d\}_{d\in\omega}$ be as in Proposition 3. Since $Y$ is $\mu_p$-random, $Y \notin \cap_d U_d$, so fix $d$ with $Y \notin U_d$. Then for all $b \geq d$, we have

$$|\overline{Y}_{N(b)} - p| < 2^{-b} \tag{*}$$

where $N(b) = 2^{3b-1}$.

If the real number $p$ is represented as a member of $2^\omega$ via

$$p = \sum_{n\in\omega} p_n 2^{-n-1} = .p_0p_1p_2\cdots$$

in binary notation, then we have to define a Turing functional $\Psi_d$ such that $p_n = \Psi_d^Y(n)$.

We pick $b \geq n+1$ such that $\overline{Y}_{N(b)} = .y_0 \cdots y_n \cdots$ is not of either of the forms

$$.y_0 \cdots y_n\, 1^{b-(n+1)} \ldots$$

$$.y_0 \cdots y_n\, 0^{b-(n+1)} \ldots$$

where as usual $1^k$ denotes a string of $k$ ones. Since $p$ is not a dyadic rational, such a $b$ exists. Then by (*) it must be that the bits $y_0 \cdots y_n$ are the first $n+1$ bits of $p$. In particular, $y_n = p_n$. So we let $\Psi_d^Y(n) = y_n$. □

## 1.4 Hippocratic results

In the last section we made it too easy for ourselves; now we will obtain the same results assuming only Hippocratic randomness.

**Theorem 5.** *There is a Hippocratic $\mu_p$-test such that if $Y$ passes this test then $Y$ computes an accumulation point $q$ of the sequence of sample averages*

$$\{\overline{Y}_n\}_{n\in\omega}.$$

*Proof.* The point is that the usual proof that each convergent sequence is Cauchy gives a $\Sigma^0_1$ class that has small $\mu_p$-measure for all $p$ simultaneously. Namely, let

$$V_d := \{Y : \exists a, b \geq d \ |\overline{Y}_{N(a)} - \overline{Y}_{N(b)}| \geq 2^{-a} + 2^{-b}\}.$$

Then $\{V_d\}_{d\in\omega}$ is uniformly $\Sigma^0_1$. Recall from Proposition 3 that we defined

$$U^p_d = \{Y : \exists b \geq d \ |\overline{Y}_{N(b)} - p| \geq 2^{-b}\}.$$

If there is a $p$ such that $|\overline{Y}_{N(b)} - p| < 2^{-b}$ for all $b \geq d$, then

$$|\overline{Y}_{N(a)} - \overline{Y}_{N(b)}| \leq |\overline{Y}_{N(a)} - p| + |p - \overline{Y}_{N(b)}| < 2^{-a} + 2^{-b}$$

for all $a, b \geq d$; thus we have

$$V_d \subseteq \cap_p U^p_d$$

and therefore

$$\mu_p(V_d) \leq \mu_p(U^p_d) \leq 2^{-d}$$

for all $p$. Thus if $Y$ is Hippocrates $\mu_p$-random then $Y \notin V_d$ for some $d$.

We next note that for any numbers $c > b$,

$$|\overline{Y}_{N(b)} - \overline{Y}_{N(c)}| < 2^{-b} + 2^{-c} < 2^{-(b-1)},$$

so $\{\overline{Y}_{N(c)}\}_{c\geq d}$ will remain within $2^{-(b-1)}$ of $\overline{Y}_{N(b)}$ for all $c > b$. That is, $\{\overline{Y}_{N(n)}\}_{n\geq d}$ is a Cauchy sequence (for each $b$ there is an $N(b)$ such that for all $n, m \geq N(b)$, $|\overline{Y}_n - \overline{Y}_m| \leq 2^{-b}$) hence $q := \lim_n \overline{Y}_{N(n)}$ exists. Write $q = .q_0q_1q_2\cdots$. Then

$$|\overline{Y}_{N(b)} - q| < 2^{-(b-1)}, \text{ so}$$

$$|\overline{Y}_{N(b+1)} - q| < 2^{-b};$$

if we define $\Theta_d$ as $\Psi_d$ in Theorem 4 except with $N(\cdot)$ replaced by $N(\cdot + 1)$, then

$$q_n = \Theta^Y_d(n).$$

and so $Y$ computes $q$ using the Turing reduction $\Theta_d$. □

To argue that the accumulation point $q$ of Theorem 5 is actually equal to $p$ under the weak assumption of Hippocratic randomness, we need:

**An analysis of the strong law of large numbers.** Let $\{X_n\}_{n\in\omega}$ be independent and identically distributed random variables with mean 0, and let $S_n = \sum_{i=0}^{n} X_i$. Then $S_n^4$ will be a linear combination (with binomial coefficients as coefficients) of the terms

$$\sum_i X_i^4,\ \sum_{i<j} X_i^3 X_j,\ \sum_{i<j<k} X_i^2 X_j X_k,\ \sum_{i<j<k<\ell} X_i X_j X_k X_\ell,\ \text{and}\ \sum_{i<j} X_i^2 X_j^2.$$

Since $\mathbb{E}(X_i) = 0$, and $\mathbb{E}(X_i^a X_j^b) = \mathbb{E}(X_i^a)\mathbb{E}(X_j^b)$ by independence, and each $X_i$ is identically distributed with $X_1$ and $X_2$, we get

$$\mathbb{E}(S_n^4) = n\,\mathbb{E}(X_1^4) + \binom{n}{2}\binom{4}{2}\,\mathbb{E}(X_1^2X_2^2)$$

$$= n\,\mathbb{E}(X_1^4) + \binom{n}{2}\binom{4}{2}\,\mathbb{E}(X_1^2)\mathbb{E}(X_2^2) = n\,\mathbb{E}(X_1^4) + \binom{n}{2}\binom{4}{2}\,\mathbb{E}(X_1^2)^2.$$

Since $0 \le \sigma^2(X_1^2) = \mathbb{E}(X_1^4) - \mathbb{E}(X_1^2)^2$, this is (writing $K := \mathbb{E}(X_1^4)$)

$$\le n\,\mathbb{E}(X_1^4) + \binom{n}{2}\binom{4}{2}\,\mathbb{E}(X_1^4) = (n + 3n(n-1))\mathbb{E}(X_1^4) = (3n^2 - 2n)K$$

so $\mathbb{E}(S_n^4/n^4) \le \frac{3K}{n^2}$. Now

$$S_n^4/n^4 \ge a^4 \cdot \mathbf{1}_{\{S_n^4/n^4 \ge a^4\}}$$

surely, so (as in the proof of Chebyshev's inequality)

$$\mathbb{E}(S_n^4/n^4) \ge a^4 \cdot \mathbb{E}(\mathbf{1}_{\{S_n^4/n^4\ge a^4\}}) = a^4 \cdot \mathbb{P}(S_n^4/n^4 \ge a^4)$$

giving

$$\mathbb{P}(\overline{X}_n = S_n/n \ge a) \le \frac{3K}{n^2a^4}$$

We now applying this to $X_n = Y_n - \mathbb{E}(Y_n) = Y_n - p$ (so that $K = K_p$). Note that (writing $\overline{p} = 1 - p$)

$$K_p = \mathbb{E}[(Y_1 - p)^4] = (1-p)^4 \cdot p + p^4 \cdot p = p\overline{p}(\overline{p}^3 + p^3) \le \frac{1}{4} \cdot 2 = \frac{1}{2},$$

so $\mathbb{P}(\exists n \ge N\ |\overline{Y}_n - p| \ge a)$ is bounded by

$$\sum_{n\ge N} \frac{3K_p}{n^2a^4} \le \frac{3}{2a^4} \sum_{n\ge N} \frac{1}{n^2} \le \frac{3}{2a^4} \int_{N-1}^{\infty} \frac{1}{x^2}\,dx = \frac{3}{2a^4(N-1)}.$$

This bound suffices to obtain our desired result:

**Theorem 6.** *If $Y$ is Hippocrates $\mu_p$-random then $Y$ satisfies the Strong Law of Large Numbers for $p$.*

*Proof.* Let $q_1$, $q_2$ be rational numbers with $q_1 < p < q_2$. Let

$$W_N := \{Y : \exists n \ge N\ \overline{Y}_n \le q_1\} \cup \{Y : \exists n \ge N\ \overline{Y}_n \ge q_2\}$$

Then $\{W_N\}_{N\in\omega}$ is uniformly $\Sigma_1^0$, and $\mu_p W_N \to 0$ effectively:

$$\mu_p\{Y : \exists n \ge N\ \overline{Y}_n \le q_1\} \le \frac{3}{2(p-q_1)^4(N-1)}$$

Thus if $Y$ is Hippocrates $\mu_p$-random then $Y \notin \cap_n W_n$, i.e., $\overline{Y}_n$ is eventually always in the interval $(q_1, q_2)$. □

**Corollary 7.** *If $Y$ is Hippocrates $\mu_p$-random then $Y$ Turing computes $p$.*

*Proof.* By Theorem 5, $Y$ computes the limit of a subsequence $\{\overline{Y}_{N(b)}\}_{b\in\omega}$. By Theorem 6, this limit must be $p$. □

Note that the randomness test in Theorem 6 depends on the pair $(q_1, q_2)$, so we actually needed infinitely many tests to guarantee that $Y$ computes $p$. This is no coincidence. Let $Y \geq_T p$ abbreviate the statement that $Y$ Turing computes $p$, i.e., $p$ is Turing reducible to $Y$.

**Theorem 8.** *For all $p$, if there is a Hippocratic $\mu_p$-test $\{U_n\}_{n\in\omega}$ such that $\{X : X \not\geq_T p\} \subseteq \cap_n U_n$, then $p$ is computable.*

*Proof.* Let $\{U_n\}_{n\in\omega}$ be such a test. By standard computability theoretic basis theorems, the complement $U_1^c$ has a low member $X_1$ and a hyperimmune-free member $X_2$. By assumption $X_1 \geq_T p$ and $X_2 \geq_T p$, so $p$ is both low and hyperimmune-free, hence by another basic result of computability theory [**?**], $p$ is computable. □

**Corollary 9.** *There is no universal Hippocratic $\mu_p$-test, unless $p$ is computable.*

*Proof.* If there is such a test then by Corollary 7 there is a test $\{U_n\}_{n\in\omega}$ as in the hypothesis of Theorem 8, whence $p$ is computable. □

# Chapter 2

# Ahlman & Koponen (2015)

**Definition 10.** Let $k \in \mathbb{N}$. An automorphisms of $\mathbb{Z}/k\mathbb{Z}$ is a bijection $f$ that preserves addition:

$$f(x+y) = f(x) + f(y)$$

**Definition 11.** The additive structure $\mathbb{Z}/k\mathbb{Z}$ is rigid if it has no nontrivial automorphisms.

**Theorem 12.** *The additive structure $\mathbb{Z}/2\mathbb{Z}$ is rigid.*

*Proof.* Otherwise $f(0) = 1$ and $f(1) = 0$, but then

$$0 = f(1) = f(1+0) = f(1) + f(0) = 0 + 1 = 1.$$

□

□

# Bibliography


[1] Itaï Ben Yaacov. Modular functionals and perturbations of Nakano spaces. *J. Log. Anal.*, 1:Paper 1, 42, 2009.

[2] Itaï Ben Yaacov and Alexander Berenstein. On perturbations of Hilbert spaces and probability algebras with a generic automorphism. *J. Log. Anal.*, 1:Paper 7, 18, 2009.

[3] Josef Berger. A decomposition of Brouwer's fan theorem. *J. Log. Anal.*, 1:Paper 6, 8, 2009.

[4] Abdelmadjid Boudaoud. Decomposition of terms in Lucas sequences. *J. Log. Anal.*, 1:Paper 4, 23, 2009.

[5] Thierry Coquand and Bas Spitters. Integrals and valuations. *J. Log. Anal.*, 1:Paper 3, 22, 2009.

[6] Isaac Goldbring. Nonstandard hulls of locally exponential Lie algebras. *J. Log. Anal.*, 1:Paper 5, 25, 2009.

[7] Karel Hrbacek. Relative set theory: internal view. *J. Log. Anal.*, 1:Paper 8, 108, 2009.

[8] Bjørn Kjos-Hanssen. Infinite subsets of random sets of integers. *Math. Res. Lett.*, 16(1):103–110, 2009.

[9] Robert S. Lubarsky and Fred Richman. Signed-bit representations of real numbers. *J. Log. Anal.*, 1:Paper 10, 18, 2009.

[10] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[11] Noopur Pathak. A computational aspect of the Lebesgue differentiation theorem. *J. Log. Anal.*, 1:Paper 9, 15, 2009.

[12] Steven Vickers. Localic completion of generalized metric spaces. II. Powerlocales. *J. Log. Anal.*, 1:Paper 11, 48, 2009.

[13] Heinz Weisshaupt. Diffusion processes via parabolic equations: an infinitesimal approach to Lindeberg's limit theorem. *J. Log. Anal.*, 1:Paper 2, 29, 2009.