Teorias da Aleatoriedade

Carlos A. P. Campani [1]
Paulo Blauth Menezes [2]

Resumo: Este trabalho apresenta uma revisão bibliográfica sobre a definição de "seqüência aleatória". Nós enfatizamos a definição de Martin-Löf e a definição baseada em incompressividade (complexidade de Kolmogorov). Complexidade de Kolmogorov é uma teoria sofisticada e profunda da informação e da aleatoriedade baseada na máquina de Turing. Estas duas definições resolvem todos os problemas das outras abordagens e satisfazem o nosso conceito intuitivo de aleatoriedade, sendo matematicamente corretas. Adicionalmente, apresentamos a abordagem de Schnorr que inclui um requisito de efetividade (computabilidade) em sua definição. São apresentadas as relações entre estas diversas definições de forma crítica.
Palavras-chave: aleatoriedade, complexidade de Kolmogorov, máquina de Turing, computabilidade, probabilidade.

Abstract: This work is a survey about the definition of "random sequence". We emphasize the definition of Martin-Löf and the definition based on incompressibility (Kolmogorov complexity). Kolmogorov complexity is a profound and sofisticated theory of information and randomness based on Turing machines. These two definitions solve all the problems of the other approaches, satisfying our intuitive concept of randomness, and both are mathematically correct. Furthermore, we show the Schnorr's approach, that includes a requisite of effectiveness (computability) in his definition. We show the relations between all definitions in a critical way.
Keywords: randomness, Kolmogorov complexity, Turing machine, computability, probability.

1 Introdução

Os matemáticos do Século XX buscavam obter uma base matemática para teoria das probabilidades. Esta base se constituiria em uma definição formal de "seqüência aleatória", que permitiria dar um significado ao cálculo de probabilidades ao estilo do proposto por Kolmogorov [20]. Esta axiomatização, proposta por Kolmogorov, embora seja bem sucedida no cálculo de probabilidades, nada afirma sobre o significado da aleatoriedade dos fenômenos físicos. A busca por tal definição é um objetivo fundamental para dar suporte a toda a

[1]Instituto de Física e Matemática, UFPel, Caixa Postal 354, CEP 96010-900, Pelotas, RS, Brazil, e Instituto de Informática, UFRGS, campani@ufpel.tche.br.
[2]Instituto de Informática, UFRGS, Caixa Postal 15064, CEP 91501-970, Porto Alegre, RS, Brazil, blauth@inf.ufrgs.br.

teoria das probabilidades e às aplicações práticas em estatística. Veremos que este esforço foi plenamente recompensado, e a meta atingida através da complexidade de Kolmogorov. Complexidade de Kolmogorov é uma teoria profunda e sofisticada da informação e da aleatoriedade baseada em teoria da computabilidade [19].

Este trabalho apresenta uma revisão bibliográfica abreviada das principais teorias desenvolvidas com este propósito. Procuramos enfatizar a teoria de Martin-Löf [28, 29] e sua relação com incompressividade (complexidade de Kolmogorov). O assunto é apresentado na ordem cronológica do desenvolvimento das diversas teorias que foram propostas, assim como procuramos mostrar as relações entre elas. É importante observar que, embora existam boas revisões bibliográficas do assunto publicadas em inglês, como [33, 34, 35, 36], e um livro bem completo [26], no Brasil nunca foi publicado um trabalho deste tipo. O texto é relativamente auto-contido, embora algum conhecimento de *teoria da medida* [13] seja desejável para o entendimento das provas de alguns teoremas (particularmente o Teorema 5). Porém, o núcleo da argumentação pode ser entendido sem este conhecimento.

Veremos que este trabalho apresenta uma (surpreendente para muitos) identificação entre *aleatoriedade* e *computabilidade*, ambas apresentadas a partir de definições matemáticas. Ou seja, veremos que uma string aleatória é aquela que não pode ser computada por uma máquina de Turing. E esta é a grande motivação do texto, ao resgatar na área de ciência da computação um problema clássico, que motivou em parte o desenvolvimento da teoria da computabilidade, e que muitas vezes passa despercebido aos pesquisadores e estudantes da área.

Além disto, embora originalmente proposta para resolver o problema de definir "aleatoriedade", a teoria apresentada nos anos sessenta, de forma independente, por Kolmogorov, Solomonoff e Chaitin [26], acabou sendo aplicada em uma vasta gama de outras aplicações e áreas tais como: inteligência artificial, complexidade computacional, biotecnologia, etc.

Assim, neste texto, estamos interessados em formalizar o conceito de seqüência aleatória. Em primeiro lugar deve-se ter clareza que a definição de aleatoriedade é dependente da distribuição de probabilidade envolvida. Por exemplo, se estivéssemos tratando com uma moeda simétrica (com iguais probabilidades de ocorrência de cara e coroa), uma seqüência em que ocorressem mais caras que coroas seria logo identificada como não aleatória. Já uma moeda em que houvesse um desequilíbrio entre suas faces, poderia resultar em uma seqüência deste tipo, a qual seria tratada como aleatória em relação a esta segunda distribuição.

Não iremos discutir aspectos físicos do lançamento da moeda, tais como aceleração, velocidade, ângulo de lançamento, etc. Nem nos preocuparemos na possibilidade de tais parâmetros serem manipulados para a obtenção de resultados "viciados". Estes fatores são irrelevantes para a nossa discussão.

Vamos considerar, na maior parte da discussão (exceção onde indicado), seqüências

obtidas pelo lançamento de uma moeda honesta (distribuição Bernoulli uniforme). “Honesta” aqui significa que a moeda é simétrica e as probabilidades de ocorrer cara ou coroa são iguais ($p = 1/2$). Representaremos esta seqüência de eventos por uma string binária infinita, em que “1” e “0” representam “cara” e “coroa”.

É óbvio que tais seqüências (seqüências infinitas) não podem ser obtidas no mundo real. Kolmogorov já havia advertido para o problema de embasar fenômenos que são essencialmente finitos através de entidades infinitas [21]. Veremos que a abordagem de Kolmogorov, baseada em incompressividade, tenta definir primeiro “aleatoriedade pontual” para então generalizar para o caso de seqüências infinitas, usando o conceito de incompressividade dos prefixos (teste seqüencial). Originalmente Kolmogorov propôs resolver o problema definindo uma seqüência aleatória como sendo aquela em que todos os prefixos são incompressíveis [22, 23]. A idéia mostrou-se errada, pois tais seqüências não existem devido à *flutuação* da complexidade [34]. O problema foi resolvido pelo uso de uma definição de complexidade que garanta que o conjunto das descrições (programas para uma máquina universal) seja *livre de prefixo*.

O Teorema 6 prova a equivalência entre a definição de seqüência aleatória de Martin-Löf, baseada em conjuntos efetivos nulos, e a definição via incompressividade usando-se complexidade de Kolmogorov. Finalmente, na Seção 6 apresentamos a definição de Schnorr, que é uma definição mais “branda” de aleatoriedade. Nas conclusões, faremos uma breve apresentação de algumas aplicações da teoria apresentada neste trabalho.

2 Kollektivs de Von Mises

Toda a teoria das probabilidades baseia-se no cálculo de novas distribuições de probabilidade a partir de outras conhecidas. Toda esta teoria deve estar baseada em uma definição matematicamente correta de “probabilidade” e “aleatoriedade”.

Muitos matemáticos entenderam que o conceito de aleatoriedade deveria ser definido como um fenômeno físico, como o é “massa”, “gravidade”, “aceleração”, etc. Ou seja, “aleatoriedade” como ela aparece nos processos físicos que ocorrem no mundo, como por exemplo o lançamento de uma moeda. Isto não quer dizer que deveremos nos aventurar em aspectos como “mecânica quântica”, fractais, etc. Mas que devemos obter uma formulação matemática que explique inteiramente o conceito. Assim, o conceito fundamental que deve ser compreendido é o de *seqüência aleatória*, já que é ela a resultante dos processos estocásticos, tais como o lançar sucessivo de uma moeda.

Von Mises foi o primeiro a propor uma solução para o problema de definir “seqüências aleatórias”. Sua definição se baseia na existência de seqüências, que ele chamou de Kollektivs, as quais possuem *limite de freqüência relativa*.

Para ilustração, sejam as seguintes strings binárias:

```
11111111111111111111
01010101010101010101
01001101011000110101
```

Determinar qual delas é mais aleatória, a partir de suas probabilidades em particular, nos conduz a um paradoxo, pois a teoria das probabilidades atribui igual probabilidade para todas (a probabilidade de ocorrer uma dada string binária de tamanho n na distribuição Bernoulli uniforme é 2^{-n}, assim, todas as três strings do exemplo tem probabilidade 2^{-20}). No entanto, duvidaríamos que as duas primeiras pudessem ter sido geradas como resultado do lançamento sucessivo de uma moeda honesta [34].

A tentativa de definir "seqüência aleatória" como sendo uma seqüência *imprevisível* pode nos conduzir a um segundo problema.

A teoria do limite de freqüência relativa interpreta uma probabilidade como uma razão entre resultados favoráveis e o total de tentativas. Tal razão deve convergir para um valor, digamos p, a medida que o número de repetições do experimento n tende a infinito. Assim, por exemplo, se estivéssemos falando de lançamentos sucessivos de uma moeda honesta, $\lim_{n\to\infty} \lambda(n)/n = 1/2$, onde $\lambda(n)$ é o número de resultados favoráveis. Dizemos que uma seqüência é *imprevisível* se um jogador ao apostar nos resultados do experimento, segundo uma regra de pagamento que considera p, não obterá melhores resultados com qualquer possível e imaginária estratégia de apostas do que obteria ao acaso (como exemplo de esquema de apostas poderíamos tomar a regra "apostar em cara após a ocorrência de três coroas seguidas"). Ou seja, o jogador não poderá ter ganhos infinitos com nenhum esquema de apostas que possa usar.

O problema aparece ao nos perguntarmos o quão longa deve ser a seqüência para garantir a convergência. Podemos dizer que a probabilidade estará na faixa de $p \pm \varepsilon$ para um n grande o suficiente, ou seja, para um $n > n_0$ (nossa definição fica dependendo de ε e n_0). Mas assim, fica evidente que nossa definição de probabilidade é uma definição circular [34].

Von Mises propôs resolver estes problemas dividindo todas as seqüências infinitas em seqüências aleatórias (que ele chamou de Kollektivs, e nós mantivemos no original neste texto), seqüências estas que possuem limite de freqüência e são apropriadas ao cálculo de probabilidades, e outras seqüências que não são de interesse da teoria das probabilidades. Ele usou evidências empíricas para postular a existência dos Kollektivs [34]. A Definição 2 apresenta a formalização da idéia de Von Mises.

Definimos $\{0,1\}^+$ como o conjunto de todas as strings binárias com um ou mais dígitos, e $\{0,1\}^\infty$ como o conjunto de todas as strings binárias infinitas unidirecionais. Se

$x \in \{0,1\}^+ \bigcup \{0,1\}^\infty$ e $x = x_0x_1x_2x_3\cdots$ então $x_0, x_1, x_2, \ldots$ são os sucessivos dígitos binários de x. Cada $x \in \{0,1\}^\infty$ determina um número real $0.x_0x_1x_2\cdots$ (pode ser representado também como $0.x$) no intervalo $[0;1) \subset \mathbb{R}$, fechado a esquerda e aberto a direita. Se x é uma string binária de tamanho maior ou igual a n então $x^n = x_0x_1x_2\cdots x_{n-1}$, ou seja, x^n denota a seqüência formada pelos n primeiros dígitos binários de x.

Definição 1 *Uma* regra de seleção de posição *é uma função parcial ϕ, que mapeia do conjunto das strings binárias finitas para os valores $\{V, F\}$, $\phi : \{0,1\}^+ \to \{V, F\}$, e que seleciona o índice $k < n$ de uma string binária $x = x_0x_1\cdots x_{n-1}$ se $\phi(x^k) = V$. A sub-seqüência obtida pela aplicação de ϕ é $x_{k_0}x_{k_1}x_{k_2}\cdots$, onde x_{k_i} é inserido na sub-seqüência se $\phi(x^{k_i}) = V$, e $\phi(x^{k+1}) = F$ para todo k tal que $k_{i-1} < k < k_i - 1$.*

Uma regra de seleção de posição é uma função parcial que seleciona uma sub-seqüência de uma seqüência dada pelo valor calculado da função (V ou F), ou seja, a função "extrai" uma nova seqüência a partir de uma seqüência dada.

A Definição 2 baseia-se claramente no problema das mesas de jogos, ao definir uma seqüência aleatória como sendo aquela cujo limite de freqüência relativa converge (ou seja, existe uma probabilidade de sucesso no jogo) e, além disto, impede que um esquema de apostas seja capaz de "quebrar a banca" (item (2) da Definição), exigindo que esta convergência seja satisfeita por qualquer sub-seqüência da seqüência dada obtida a partir dela pela aplicação de uma função (esquema de aposta).

Definição 2 *Uma seqüência binária infinita $x = x_0x_1x_2\cdots$, é uma seqüência aleatória (Kollektiv) se:*

1. $\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} x_i}{n} = p$;
2. *Toda seqüência $x_{k_0}x_{k_1}x_{k_2}\cdots$, obtida de x pela aplicação de alguma regra de seleção de posição* admissível *ϕ, deve satisfazer o item 1 desta Definição.*

Isto significa que a seqüência infinita original e todas as possíveis sub-seqüências obtidas por alguma regra de seleção de posição admissível devem possuir um mesmo limite de freqüência. É importante observar que a Definição 2 não define o que é uma regra de seleção de posição *admissível*.

Existe evidência empírica que o item (1) da Definição 2 é satisfeito. Mas nenhuma evidência empírica pode garantir a convergência. Eventualmente, a seqüência diverge a partir de algum ponto. O problema é que todo experimento no mundo real é finito, e a evidência empírica não pode garantir uma definição que exige uma seqüência infinita. Quanto ao

item (2), ele garante que nenhum esquema de apostas, usando alguma função parcial como regra de seleção de posição admissível, pode selecionar alguma sub-seqüência em que seja possível obter ganhos infinitos.

Um problema que aparece é o uso de funções parciais como regras de seleção de posição. Seja, por exemplo, a função parcial $\phi_0(x^i) = V$ se $x_i = 1$ e indefinido caso contrário [34]. Então, para a sub-seqüência selecionada, $p = 1$. Se permitirmos funções como ϕ_0 então não existem Kollektivs.

3 Funções de Seleção de Wald-Church

Para viabilizar a proposta de Von Mises, Wald propôs tomar apenas um conjunto contável de funções para ser usado como regras de seleção de posição. Fazendo isto, os Kollektivs passam a existir e a Definição 2 faz sentido [34].

Church propôs que fossem tomadas como regras de seleção de posição apenas as funções parciais recursivas. Isto significa que uma seqüência aleatória seria imune apenas a toda tentativa de ganho infinito baseada em uma estratégia *computável* (sobre computabilidade veja [19]).

Assim, podemos reescrever a Definição 2 restringindo ϕ apenas às funções parciais recursivas. Estas seqüências são chamadas *Wald-Church aleatórias*.

Ainda existe um problema com esta definição: Existem seqüências Wald-Church aleatórias x, com limite de freqüência de $1/2$, mas que satisfazem $\sum_{i=0}^{n-1} x_i/n \geq 1/2$, para todo n [34]. Estas seqüências permitem ganhos infinitos se o jogador apostar no "1", o que obviamente invalida a seqüência de ser aleatória num sentido mais amplo.

4 Definição de Martin-Löf

A definição de seqüência aleatória proposta por Martin-Löf [28, 29], isenta das dificuldades das outras abordagens apresentadas nas duas Seções precedentes, é baseada em *conjuntos efetivos nulos* e na existência de um *conjunto efetivo nulo maximal* (veja também [33, 34, 36]). Ela é uma abordagem *quantitativa* e baseada em teoria da medida [13, 16].

Teoria da medida tem seus fundamentos nas áreas de análise, Calculus e espaços métricos, e foi desenvolvida como uma generalização do cálculo de áreas e volumes. Suas principais aplicações estão no cálculo de integrais, para os casos nos quais a integral de Riemann não pode ser aplicada, e na teoria das probabilidades, ao permitir o cálculo de probabilidades em espaços "exóticos". A idéia de "medida" fica bem compreendida se pensarmos em medir o tamanho de conjuntos, mesmo que eles sejam não contáveis ou "mal comportados", como

é o caso, por exemplo, do conjunto de Cantor [27]. Uma medida pode ser entendida como uma probabilidade (neste caso é chamada de *medida de probabilidade*).

Na abordagem de Martin-Löf isto é importante pois ele identifica "aleatoriedade" com "propriedade de ser típico" (ou seja, com alta probabilidade de ocorrência). Assim, a "medição" do tamanho de conjuntos torna-se central, e determina se os elementos de um dado conjunto são aleatórios ou não.

Considere o conjunto de todas as seqüências binárias infinitas geradas pelo lançamento de uma moeda honesta. Intuitivamente nós chamamos uma seqüência "aleatória" se ela é *típica* [33], porque ela é representativa desta *classe* de seqüências. "Típico" aqui significa "desprovido de características peculiares". Se ela não é típica, então ela é "especial", pois possui alguma propriedade que a distingue das demais (por exemplo, uma seqüência em que cada coroa é seguida por uma cara). As seqüências especiais formam um conjunto com medida zero (significando que tem baixa probabilidade de ocorrerem).

Temos que ter cuidado com o que chamamos de "especial". Se a propriedade "ser aleatório" é considerada como "especial", então obtemos um evidente paradoxo.

Paradoxos (ou antinômias) são bastante freqüentes na história da matemática. Podemos citar o famoso "paradoxo do barbeiro", de autoria de Bertrand Russel, que diz que "o barbeiro é aquele que barbeia os homens que não se barbeiam a si próprios". O paradoxo é obtido ao perguntar se o barbeiro se barbeia a si próprio [30]. Obteriamos um paradoxo semelhante ao identificar aleatoriedade com "ser especial".

Entendemos como típico um indivíduo que é representativo de sua classe. Por exemplo, se dissermos que João Vitor é um típico menino de três anos de idade, queremos dizer que além de João Vitor possuir três anos de idade, ele também possui todas as características específicas que o distinguem como pertencente à classe dos meninos de três anos de idade. Ou seja, ele possui todas as propriedades que são compartilhadas por todos os meninos de três anos de idade.

Podemos dizer que as seqüências típicas são aquelas que respeitam todas as leis da aleatoriedade (estocasticamente falando), tal como a Lei dos Grandes Números [16, 17]. Cada seqüência satisfazendo uma destas propriedades pertence a uma "maioria". Assim, uma seqüência típica pertence a todas as maiorias, ou satisfaz todas as leis de aleatoriedade. Logo, o conjunto das seqüências aleatórias seria a intersecção de todas estas maiorias.

Isto nos conduz a um problema: Cada seqüência x em particular induz um teste de aleatoriedade, $\delta_x = \{0,1\}^\infty - \{x\}$, que a exclui de uma maioria. Assim, a intersecção de todas as maiorias seria vazia, ou seja, não existiriam seqüências aleatórias!

De outra forma podemos dizer que um elemento $x \in \Omega$ é típico se qualquer subconjunto $U \in \Omega$ que contém uma parte pequena de Ω, não contém x. Mas isto nos conduz a dizer

que qualquer x não é aleatório pois $\{x\}$ contém x e é uma parte pequena de Ω. Assim, retornando ao nosso exemplo com crianças de três anos de idade, cada uma pode ser considerada um indivíduo atípico, porque seus gostos pessoais específicos forma uma classe pequena do conjunto de crianças [33].

Martin-Löf resolveu este problema restringindo as leis de aleatoriedade (testes de aleatoriedade) apenas aos testes efetivos (computáveis). Isto significa que apenas "regularidades" computáveis serão testadas, nada interessando eventuais "regularidades" não computáveis, por não serem mecanicamente detectáveis.

A ocorrência de uma seqüência "regular" no lançamento de uma moeda honesta é um acontecimento "especial" ou "notável". Por exemplo, se a seqüência possuísse um "1" após a ocorrência de dois "0" consecutivos ela seria "especial", e a sua propriedade poderia ser facilmente testada. Ela não possuiria as propriedades estocásticas necessárias para ser considerada aleatória.

Aqui entendemos "regularidade" no sentido da seqüência possuir alguma propriedade estocástica que a exclui do conjunto das seqüências aleatórias. Martin-Löf propõe que estas propriedades devem ser testadas usando-se alguma função parcial recursiva (teste de aleatoriedade), que exclui aquelas consideradas "especiais". Assim, a definição de Martin-Löf propõe a existência de um conjunto de seqüências, ditas aleatórias, que possui complemento recursivamente enumerável.

Seja o conjunto $X \subset \{0,1\}^\infty$. Dizemos que X é um *conjunto nulo* se $\Pr(X)$ é definido e $\Pr(X) = 0$. Uma outra forma de definir conjunto nulo é apresentada na Definição 3.

Denotamos o tamanho da string binária x (número de dígitos binários de x) como $|x|$. Chamamos x *prefixo* de z se $z = xy$ (z é a concatenação de x e y). Definimos o *cilindro* Γ_x denotando o conjunto $\{a \in \{0,1\}^\infty, |x| = n : a^n = x\}$, ou seja, o conjunto de todas as strings binárias infinitas que tem como prefixo a^n. Um cilindro define geometricamente um sub-intervalo no intervalo $[0;1) \subset \mathbb{R}$. Assim, por exemplo, Γ_x pode ser associado com o intervalo $[0.x; 0.x + 2^{-|x|})$ [26]. O cilindro Γ define uma *medida* trivial (dada pelo tamanho dos intervalos) no intervalo $[0;1)$.

Definição 3 *$X \subset \{0,1\}^\infty$ é um* conjunto nulo *se para todo $\varepsilon > 0$ existe uma seqüência de strings binárias $x_0, x_1, x_2 \ldots$ tal que:*

1. $X \subset \Gamma_{x_0} \bigcup \Gamma_{x_1} \bigcup \Gamma_{x_2} \bigcup \cdots$;
2. $\sum_{i=0}^{\infty} 2^{-|x_i|} < \varepsilon$.

Na Definição, $\bigcup_{i=0}^{\infty} \Gamma_{x_i}$ é chamado de *cobertura* de X. Os intervalos Γ_x formam uma

topologia [27] em $[0;1)$ e representam subconjuntos de Borel em $[0;1)$ [13]. A Definição 4 estende a anterior adicionando um requisito de efetividade.

Definição 4 $X \subset \{0,1\}^\infty$ *é um* conjunto efetivo nulo *se existe um algoritmo que recebe um número racional* $\varepsilon > 0$ *como entrada e enumera o conjunto de strings* $\{x_0, x_1, x_2 \ldots\}$ *tal que:*

1. $X \subset \Gamma_{x_0} \bigcup \Gamma_{x_1} \bigcup \Gamma_{x_2} \bigcup \cdots$;
2. $\sum_{i=0}^{\infty} 2^{-|x_i|} < \varepsilon$.

Observe que podemos substituir, na Definição 4, ε por 2^{-m}, $m > 0$, e $<$ por $\leq$ em (2), sem perda de generalidade.

A Definição 4 diz que um conjunto efetivo nulo é um conjunto com medida zero (dado pelo ε que pode ser tão pequeno quanto se queira) que pode ser efetivamente (algoritmicamente) gerado. Ou seja, se o conjunto é nulo e recursivamente enumerável.

Qualquer subconjunto de um conjunto efetivo nulo é também um conjunto efetivo nulo. Um conjunto unitário $\{x\}$ é um conjunto nulo se x é computável (ou não aleatório).

Por "algoritmo de cobertura" para um conjunto efetivo nulo nós entendemos o algoritmo mencionado na Definição 4.

Lema 1 *Seja* $X_0, X_1, X_2, \ldots$ *uma seqüência de conjuntos efetivos nulos tal que existe um algoritmo que, para todo* $i \geq 0$ *e* $\varepsilon > 0$*, produz algum algoritmo de cobertura para* X_i*. Então,* $\bigcup_{i=0}^{\infty} X_i$ *é um conjunto efetivo nulo.*

Prova: Para obter uma ε*-cobertura de* $\bigcup X_i$*, nós obtemos uma* $(\varepsilon/2)$*-cobertura de* X_0*, uma* $(\varepsilon/4)$*-cobertura de* X_1*, uma* $(\varepsilon/8)$*-cobertura de* X_2*, e assim por diante. Para gerar esta cobertura combinada, nós usamos o algoritmo, dado no enunciado, que produz coberturas para* X_i *a partir de* i*.*

O Lema 1 prova que a união *contável* (ou seja, em que o número de operações de união é equipotente ao conjunto dos números naturais) de conjuntos efetivos nulos também é um conjunto efetivo nulo.

O Teorema 1 prova que a união de todos os conjuntos efetivos nulos é um conjunto efetivo nulo. Ou seja, prova a existência de um *conjunto efetivo nulo maximal* [28, 29, 33].

Teorema 1 *Existe um* conjunto efetivo nulo maximal*, isto é, um conjunto efetivo nulo* M *tal que* $X \subset M$ *para todo conjunto efetivo nulo* X*.*

Prova: Não podemos usar o Lema 1 para todos os conjuntos efetivos nulos pois o Lema se aplica a um conjunto contável deles, e o conjunto de todos os conjuntos efetivos nulos não é contável (já que qualquer subconjunto de um conjunto efetivo nulo também é um conjunto efetivo nulo). Devemos, ao contrário, considerar todos os algoritmos de cobertura (já que o conjunto dos algoritmos é contável).

Para um dado algoritmo, que a partir de um número racional positivo gera strings binárias, não podemos determinar efetivamente se ele é um algoritmo de cobertura ou não. No entanto, podemos faze-lo por aproximação. Se um algoritmo, para um dado $\varepsilon > 0$, gera strings $x_0, x_1, x_2, \ldots$, nós podemos verificar se $2^{-|x_0|} + \cdots + 2^{-|x_k|} < \varepsilon$ ou não. Se não, nós apagamos x_k da seqüência gerada. Seja A' o algoritmo modificado obtido a partir de A. Sabemos que:

1. *Se A é um algoritmo de cobertura de algum conjunto efetivo nulo, então A' é idêntico a A (a condição nunca é violada);*
2. *Para qualquer algoritmo A, A' é um algoritmo de cobertura de algum conjunto efetivo nulo.*

Podemos usar o mesmo argumento do Lema para todos os algoritmos $A'_0, A'_1, A'_2, \ldots$, onde $A_0, A_1, A_2, \ldots$ é a seqüência de todos os algoritmos que recebem como entrada um racional positivo e geram strings binárias.

Assim, o conjunto efetivo nulo maximal contém todos os conjuntos efetivos nulos. Isto equivale a afirmar a existência de um *teste de aleatoriedade universal*, capaz de encontrar toda e qualquer regularidade (computável) em uma seqüência qualquer. Resolve-se assim os problemas que as outras abordagens para definição de aleatoriedade tinham, pois este conjunto maximal contém realmente *todas* as seqüências não aleatórias.

Definição 5 *Uma seqüência x de zeros e uns é chamada* (Martin-Löf) aleatória *com respeito a medida de Bernoulli uniforme se x não pertence ao conjunto efetivo nulo maximal.*

Uma outra forma de dizer isto é afirmar que x não pertence a nenhum conjunto efetivo nulo.

As Definições 6 e 7 apresentam a idéia de *testes de aleatoriedade* e generalizam a definição de aleatoriedade apresentada até aqui (baseada em distribuição Bernoulli uniforme) para distribuições de probabilidade quaisquer. Importante afirmar que, no caso de seqüências finitas, só podemos falar em *graus de aleatoriedade* (e por isto a existência dos *níveis de significância* nas definições que se seguem), pois tal conceito só se torna absoluto quando aplicado a seqüências infinitas.

Definição 6 *Seja P uma distribuição de probabilidade recursiva sobre o espaço amostral S. Uma função $\delta : S \to \mathbb{N}$ é um P-*teste *(*teste de Martin-Löf*) se:*

1. *δ é enumerável (o conjunto $V = \{(m; x) : \delta(x) \geq m\}$ é recursivamente enumerável);*
2. *$\sum\{P(x) : \delta(x) \geq m, |x| = n\} \leq 2^{-m}$, para todo n.*

A função δ testa *deficiência de aleatoriedade*. Assim, a expressão $\delta(x) \geq m$ *rejeita* x com nível de significância 2^{-m}. As *regiões críticas* associadas ao teste são os conjuntos $V_m = \{x : \delta(x) \geq m\}$ [28, 34].

Observe que as regiões críticas são aninhadas, ou seja, $V_m \supseteq V_{m+1}$, para $m \geq 1$. O complemento da região crítica V_m é chamado de intervalo de confiança $(1 - 2^{-m})$. Se $x \in V_m$ então "x é aleatório" é rejeitado com nível de significância 2^{-m}.

A definição de Martin-Löf é adequada no sentido que para cada conjunto nulo X, a propriedade "não pertence à X" vale para todas as seqüências aleatórias.

Devemos, a partir da definição de *aleatoriedade pontual* dada na Definição 6, generalizar para o caso de seqüências infinitas, através do conceito de *teste seqüencial*, que significa aplicar a definição anterior aos prefixos da seqüência aleatória e tomar o supremo.

A Definição 7 introduz esta idéia de *teste seqüencial*. Nela, $\mu(.)$ é uma *medida de probabilidade* no espaço amostral $\{0, 1\}^\infty$ com relação à σ-álgebra dos intervalos em $[0; 1)$. $\mu(.)$ é simplesmente uma *medida de probabilidade* que generaliza a medida trivial que usamos até agora, e a σ-álgebra é uma forma matemática de definir um espaço amostral, sobre o qual se aplica a medida, de forma mais genérica. Mais informações sobre teoria da medida podem ser encontradas em [13].

Definição 7 *Seja μ uma medida de probabilidade recursiva sobre o espaço amostral $\{0, 1\}^\infty$. Uma função total $\delta : \{0, 1\}^\infty \to \mathbb{N} \cup \{\infty\}$ é um μ-teste seqüencial (μ-teste de aleatoriedade seqüencial de Martin-Löf) se:*

1. *$\delta(x) = \sup_{n\in\mathbb{N}}\{\gamma(x^n)\}$, onde $\gamma : \{0, 1\}^* \to \mathbb{N}$ é uma função enumerável total ($V = \{(m; x) : \gamma(x) \geq m\}$ é um conjunto recursivamente enumerável);*
2. *$\mu\{x : \delta(x) \geq m\} \leq 2^{-m}$, para cada $m \geq 0$.*

O item 1 é aquele que se aplica sobre os prefixos de x (x^n) e toma o supremo (que é o limite do valor de $\gamma(x^n)$). Já o item 2 afirma simplesmente que os conjuntos com menor medida são aqueles que excedem um determinado valor m tomado como nível de significância ($\delta(x) \geq m$ e o nível de significância é 2^{-m}).

Se $\delta(x) = \infty$ então dizemos que x falha para δ, ou que δ rejeita x. O conjunto dos x rejeitados por δ tem, por definição, μ-medida zero (baixa probabilidade de ocorrerem em um processo estocástico). Ou seja, estes rejeitados são considerados "não típicos" ou "não aleatórios".

5 Complexidade de Kolmogorov e Incompressividade

A complexidade de Kolmogorov foi proposta como uma teoria da informação baseada no tamanho do menor programa para uma máquina universal [12] que computa uma determinada saída (string binária) [22, 23, 26]. Kolmogorov pretendia definir uma string binária aleatória como sendo aquela cuja menor descrição não é muito menor que a própria string (este conceito chama-se de *incompressividade*).

A idéia original de Kolmogorov tornou-se possível quando foram admitidas apenas descrições *livres de prefixo*. Esta exigência está muito relacionada com a desigualdade de Kraft (veja Teorema 3), expressão muito conhecida na teoria da informação clássica. Isto significa simplesmente que basta os programas conhecerem seu próprio tamanho. Isto não é uma exigência absurda, já que a maioria das linguagens de programação possui algum comando que indica o fim do programa (como por exemplo o `END.` do PASCAL).

Esta idéia de "tamanho do menor programa" está associada à idéia de "caótico", no sentido de "desordem" ou "entropia" (veja [11]), embora pareça contraditório estarmos tentando definir formalmente algo que dizemos "desordenado", pois tal definição representaria uma "ordenação". Esta idéia de aleatoriedade associada ao tamanho da descrição das strings pode ser entendida simplesmente ao confrontarmos a possibilidade de definir o número 1.000.000.000, na base 10, simplesmente como 10^9, enquanto teriamos dificuldade de fazer o mesmo com o número 5.359.871.331, formulado ao acaso. Podemos intuir daí uma importante propriedade do conjunto de todas as seqüências: a esmagadora maioria das seqüências são irregulares, simplesmente, por um argumento de contagem, porque faltam descrições curtas suficientes para descreve-las.

É importante observar que esta definição, sendo bem sucedida em definir "seqüências aleatórias", vincula o conceito formal de "computabilidade", como definido por Turing, Church e outros [19], com o conceito de "aleatoriedade", como aplicado na área de probabilidade e estatística.

Um conjunto $\beta \subseteq \{0,1\}^+$ é *livre de prefixo* se nenhum elemento de β é prefixo de outro elemento de β. Uma *codificação binária* $E : \{0,1\}^+ \rightarrow \{0,1\}^+$ é um mapeamento sobre o conjunto de todas as strings binárias de tamanho maior que zero. O mapeamento E associa uma *palavra de código* $E(x)$ para cada string x. Se o conjunto de códigos gerado pela codificação E é livre de prefixo, chamamos E de *codificação livre de prefixo*.

Consideremos uma máquina de Turing $\mathcal{M}$ que a partir de uma string binária p computa a saída x, $\mathcal{M}_p = x$. Nós dizemos que $\mathcal{M}$ interpreta p como uma descrição de x. Em outras palavras, p é um $\mathcal{M}$-programa que produz x. Supomos que se $\mathcal{M}_p$ está definida então $\mathcal{M}_q$ não está definida para nenhum prefixo q de p. Obtemos isto através do uso de codificação livre de prefixo e chamamos estes programas de *programas auto-delimitados*. A máquina $\mathcal{M}$ que recebe como entrada programas auto-delimitados é chamada de máquina de Turing de prefixo.

Definição 8 *Seja $\mathcal{M}$ uma máquina de Turing de prefixo. A* complexidade $K_{\mathcal{M}}(x)$ *de uma string binária x é o comprimento da menor descrição p, usando-se codificação livre de prefixo, tal que $\mathcal{M}_p = x$,*

$$K_{\mathcal{M}}(x) = \min_{\mathcal{M}_p = x} |p| \ .$$

Se não existe uma descrição p de x então dizemos, por definição, que $K_{\mathcal{M}}(x) = \infty$.

Resta provar que a complexidade definida desta maneira é invariante com relação à máquina $\mathcal{M}$ escolhida para expressa-la, permitindo reescrever $K_{\mathcal{M}}(x)$ como $K(x)$. Isto será feito no Teorema 2 usando-se a propriedade da máquina universal de simular qualquer outra máquina na enumeração $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \ldots$ das máquinas de Turing (ou dos programas para a máquina de Turing).

Denotamos a seqüência de n repetições de um símbolo a como a^{*n}. Nós codificaremos as máquinas da enumeração mencionada como

$$\overbrace{111\cdots1}^{i \text{ vezes}}0p = 1^{*i}0p\,,$$

significando que a máquina universal espera encontrar concatenado à esquerda da descrição p uma descrição de qual máquina irá simular. Ou seja, a máquina universal $\mathcal{U}$ irá simular a execução do programa p em $\mathcal{M}_i$, a i-ésima máquina da enumeração padrão.

Teorema 2 *(Teorema da Invariância) Existe uma máquina $\mathcal{U}$, chamada* universal*, tal que para qualquer máquina $\mathcal{M}$ e string binária x, e para algum $c > 0$, $K_{\mathcal{U}}(x) \leq K_{\mathcal{M}}(x) + c$ e c depende de $\mathcal{M}$, mas não de x.*

Prova: Para $K_{\mathcal{M}}(x) = \infty$ a desigualdade é trivialmente verdadeira. Podemos mostrar uma máquina universal simulando outra máquina qualquer. Por exemplo, considere a máquina universal $\mathcal{U}$ tal que $\mathcal{U}_{1^{*i}0p} = \mathcal{M}_{i_p}$, onde $\mathcal{M}_i$ é a i-ésima máquina na enumeração das máquinas. Suponha que $\mathcal{M}$ é a n-ésima máquina na enumeração, ou seja, $\mathcal{M}_n = \mathcal{M}$. Logo,

$$K_{\mathcal{M}}(x) = \min_{\mathcal{M}_p = x} |p|$$

e

$$K_{\mathcal{U}}(x) = \min_{\mathcal{U}_{1^{*n}0p}=x} |1^{*n}0p| = \min_{\mathcal{U}_{1^{*n}0p}=x} |p| + n + 1 = K_{\mathcal{M}}(x) + n + 1\,.$$

Ou seja, o limite superior da complexidade expressa em $\mathcal{U}$ é $K_{\mathcal{M}}(x)+n+1$ e eventualmente existe um p' tal que $\mathcal{U}_{p'} = x$ e $|p'| < |p|+n+1$. Assim, $K_{\mathcal{U}}(x) \leq K_{\mathcal{M}}(x)+n+1$. Tomando $c = n+1$ prova-se o Teorema, com c dependendo apenas de $\mathcal{M}$.

Definição 9 *Se $K(x) \geq |x| - c$, para algum $c > 0$, dizemos que x é c-*incompressível*. Se existe um c tal que x é c-incompressível, então dizemos que x é* incompressível*.*

Tomemos como exemplo para ilustrar a Definição 9 as seqüências apresentadas no início da Seção 2. A primeira seqüência, `11111111111111111111`, que tem tamanho vinte bits, poderia ser computada pelo programa:

```
FOR I=1 TO 20 PRINT 1
```

cujo tamanho é $\log 20 + O(1)$, onde $\log(.)$ é o logaritmo na base 2 e $O(1)$ é a notação assintótica para constante. Já por sua vez a terceira seqüência, `01001101011000110101`, teria que ser computada por um programa que tivesse a mesma armazenada literalmente no próprio programa:

```
PRINT 01001101011000110101
```

Resulta que a segunda string, diferentemente da primeira, não poderia ser computada por um programa que fosse significativamente menor que 20, o tamanho da própria string, sendo esta segunda string, segundo a definição, incompressível.

Isto indica a possibilidade de definirmos uma seqüência aleatória infinita x como sendo aquela em que, para algum $c > 0$, $K(x^n) \geq n - c$, para todo n, vinculando, desta forma, o conceito de incompressividade ao de aleatoriedade. Provaremos mais adiante que esta definição é matematicamente equivalente à definição de Martin-Löf.

O Teorema 3 é um resultado clássico e bem conhecido de teoria da informação e será usado para provar uma propriedade importante de $K(.)$.

Teorema 3 *(Desigualdade de Kraft) Toda codificação livre de prefixo E satisfaz a desigualdade $\sum_{x\in\alpha} 2^{-|x|} \leq 1$, onde α é o conjunto de códigos gerados por E.*

Prova: Mostramos na Figura 1 uma árvore que ilustra todos os possíveis caminhos para obter uma string binária. A árvore segue infinitamente para baixo. Cada caminho

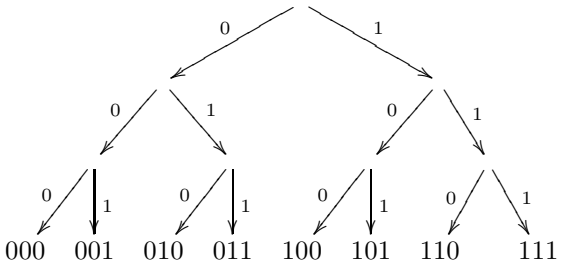

Figura 1. Árvore da codificação binária

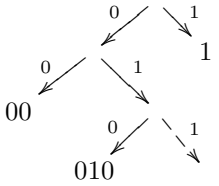

Figura 2. Árvore mostrando códigos livres de prefixo

representa uma string binária. Podemos ver na Figura 2 os pontos marcados indicando três códigos livres de prefixo (estes códigos foram tomados como exemplo, não tendo influência no resutado da prova do Teorema), desde que nos caminhos abaixo dos pontos não exista nenhum outro código. São eles 00, 010 e 1.

Para alocar um código na árvore devemos caminhar pela árvore começando pela raiz. Uma vez chegando a um ponto de bifurcação na árvore, devemos tomar uma de duas ações:

- *Neste ponto já foi encontrado um código, então pare;*
- *Ainda não é um código, então mova-se mais um passo para baixo com* $\Pr(0) = \Pr(1) = \frac{1}{2}$.

Após caminhar n *passos na árvore nós podemos ter terminado em algum nível* $m < n$ *se encontramos um código (pois os códigos são livres de prefixo), ou podemos ter atingido o*

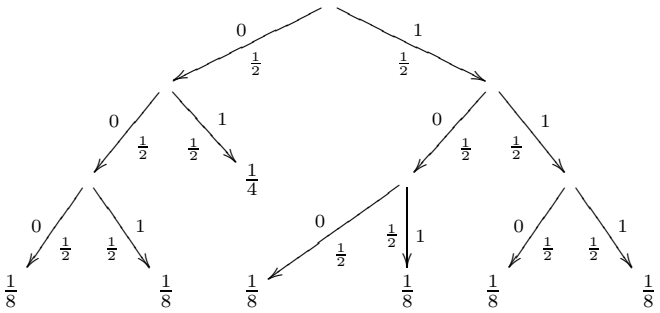

Figura 3. Árvore mostrando a conservação da probabilidade

nível n sem encontrar um código. De qualquer forma a probabilidade de qualquer caminho na árvore pode ser calculada como:

- *Caminho com n bits (não encontrou código):* $\Pr(x^n) = \left(\frac{1}{2}\right)^n$*;*
- *Caminho com $m < n$ bits (encontrou código):* $\Pr(x^m) = \left(\frac{1}{2}\right)^m$.

Pela propriedade da conservação da probabilidade *a soma das probabilidades de qualquer nível da árvore deve ser igual a 1. Ilustramos esta propriedade na Figura 3. Nesta figura apresentamos uma árvore hipotética em que o desenvolvimento de três níveis da árvore resultou em uma codificação encontrada no nível 2 (código 01). Podemos somar as probabilidades obtendo* $1/8 + 1/8 + 1/4 + 1/8 + 1/8 + 1/8 + 1/8 = 1$.

Desta forma podemos dizer que

$$\sum_{\text{códigos com } |x|<n} \left(\frac{1}{2}\right)^{|x|} + \sum_{\text{caminhos ainda não terminados}} \left(\frac{1}{2}\right)^{n} = 1\,.$$

Como

$$\sum_{\text{caminhos ainda não terminados}} \left(\frac{1}{2}\right)^{n} \geq 0\,,$$

segue que

$$\sum_{\text{códigos com } |x|<n} \left(\frac{1}{2}\right)^{|x|} \leq 1\,.$$

Tomando o $\lim_{n\to\infty}$ *obtemos*

$$\lim_{n\to\infty} \sum_{\text{códigos com } |x|<n} \left(\frac{1}{2}\right)^{|x|} = \sum_{x\in\alpha} \left(\frac{1}{2}\right)^{|x|},$$

e está completa a prova.

Teorema 4

$$\sum_{x\in\{0,1\}^+} 2^{-K(x)} \leq 1$$

Prova: Trivial. Conseqüência da desigualdade de Kraft.

O Teorema 5 apresenta um resultado profundo de complexidade de Kolmogorov que será importante para a argumentação que se seguirá (veja mais sobre este resultado em [8, 14, 26]) . Nele afirmamos a existência de uma semi-medida enumerável discreta universal.

A teoria das semi-medidas estende a teoria da medida [13] (veja uma revisão de teoria das semi-medidas em [26]). Uma semi-medida enumerável é uma semi-medida cuja função (real ou discreta) não é computável (função parcial recursiva, como definida por Turing, Church e outros), mas pode ser aproximada por cima ou por baixo, com a precisão que se queira. Tal definição é bem menos restritiva que a exigência de ser computável, não tornando assim o conjunto das funções sobre as quais se aplica tão restrito, como poderia parecer a princípio.

Teorema 5 $K(x) \leq -\log\mu(x) + O(1)$, *para qualquer semi-medida enumerável discreta* $\mu(.)$, *onde* $\log(.)$ *denota o logaritmo na base 2, e* $O(1)$ *é a notação assintótica denotando um termo constante.*

Ou seja, o Teorema afirma que $2^{-K(x)}$ é uma semi-medida *universal*, que pode substituir qualquer outra semi-medida $\mu(x)$ (por *dominação* [26]).

O Teorema 6 coroa o presente trabalho fornecendo uma prova da equivalência entre a definição de seqüência aleatória de Martin-Löf e a definição via incompressividade.

Teorema 6 *Uma seqüência* $x = x_0x_1x_2x_3\cdots$ *é* Martin-Löf aleatória *se e somente se, para alguma constante* $c > 0$, $K(x^n) \geq n - c$, *para todo* n.

Prova: (Se) Basta provar que o conjunto N *das seqüências* x *com* $K(x^n) < n - c$, *para qualquer* n, *é um conjunto efetivo nulo. Devemos observar que:*

1. N *é enumerável;*
2. $\sum_{x \in N} 2^{-|x|} < 2^{-c}$.

A primeira afirmação é trivial pois $K(.)$ *é co-enumerável. A segunda afirmação pode ser facilmente provada pois:*

$$\begin{aligned} K(x) &< |x| - c \\ -|x| &< -K(x) - c \\ 2^{-|x|} &< 2^{-K(x)} 2^{-c} \\ \sum_x 2^{-|x|} &< 2^{-c} \sum_x 2^{-K(x)} \qquad (1) \\ \sum_x 2^{-|x|} &< 2^{-c} \qquad (2) \end{aligned}$$

A passagem de (1) para (2) é devido ao fato que $\sum_x 2^{-K(x)}$*, pela desigualdade de Kraft, é uma série que converge (veja Teorema 4).*

(Somente se) Observe que podemos definir uma função computável f *que recebe um valor* $c > 0$ *qualquer e um inteiro* $i = 0, 1, 2, \ldots$ *e gera a cobertura* $\Gamma_{f(c,0)}, \Gamma_{f(c,1)}, \Gamma_{f(c,2)}, \ldots$*, que cobre* N *e tem medida no máximo* 2^{-2c}*. Observe que esta seqüência existe e pode ser efetivamente gerada.*

Considere a função $g(c, i) = |f(c, i)| - c$*. Assim, para um dado* c*,*

$$\sum_i 2^{-g(c,i)} = \sum_i 2^{-|f(c,i)|+c} = \sum_i 2^c 2^{-|f(c,i)|} = 2^c \sum_i 2^{-|f(c,i)|} \leq 2^c 2^{-2c} = 2^{-c}\,.$$

Portanto, a soma $\sum_{c,i} 2^{-g(c,i)}$*, sobre todos os* c *e* i*, não excede 1, pois esta série é obviamente convergente. Isto garante que* $\sum 2^{-g(c,i)}$ *é uma semi-medida enumerável, já que* $g(c, i)$ *é computável, desde que evitemos que ocorram coincidências do valor de* $f(c, i)$ *para diferentes pares* (c, i)*. Então, considere a semi-medida*

$$\mu(x) = \sum \{2^{-g(c,i)} | f(c, i) = x\}\,.$$

$\mu(.)$ *é enumerável por baixo, porque* f *e* g *são computáveis.*

Sabemos que $K(x) \leq -\log \mu(x) + O(1)$ *(Teorema 5) e, fixando um par* (c, i) *obtemos*

$$K(f(c, i)) \leq g(c, i) + O(1) = |f(c, i)| - c + O(1)\,.$$

Observe que $f(c, i)$ *são os prefixos das seqüências que pertencem a* N*. Isto completa a prova do Teorema.*

6 Definição de Schnorr

Schnorr propôs uma modificação na definição de Martin-Löf, pela adição de um requisito a mais [33]. Esta nova definição estabelece um conjunto de strings típicas que é maior que o da definição de Martin-Löf.

Nesta nova abordagem, um conjunto $X \subset \{0,1\}^\infty$ é chamado *conjunto efetivo nulo de Schnorr* se existe uma função $f(\varepsilon, i)$, definida para todos os números racionais $\varepsilon > 0$ e naturais $i \geq 0$, e uma função *computável* $g(\varepsilon, \eta)$, definida sobre todos os racionais $\varepsilon, \eta > 0$, tal que:

1. $X \subset \bigcup_{i=0}^{\infty} \Gamma_{f(\varepsilon,i)}$ para todo $\varepsilon > 0$;
2. $\sum_{i=0}^{\infty} \mu(\Gamma_{f(\varepsilon,i)}) < \varepsilon$ para todo $\varepsilon > 0$;
3. $\sum_{i>g(\varepsilon,\eta)} \mu(\Gamma_{f(\varepsilon,i)}) < \eta$ para todo $\varepsilon, \eta > 0$.

A condição adicional (3) implica que a série $\sum_i \mu(\Gamma_{f(\varepsilon,i)})$ convergirá para um valor menor que ε (pois existem menos termos a serem somados que satisfazem (3)) e que, além disto, *convergirá de forma efetiva*.

Convergir de forma efetiva significa que para cada $\eta > 0$ podemos encontrar de forma algorítmica um valor que não difere da soma da série mais que η. Isto significa que esta série pode ser aproximada de forma computável. Para isto é necessário que as funções f e g sejam totais (requisito que inexistia na definição de Martin-Löf).

A versão de Schnorr da definição de seqüência típica implica que:

1. Para a distribuição Bernoulli uniforme (e para todas as distribuições usuais) não existe um conjunto nulo maximal. Basta observar que para todo conjunto efetivo nulo de Schnorr existe uma seqüência computável que não pertence ao conjunto. Assim, o conjunto $\{x\}$ é um conjunto efetivo nulo de Schnorr para cada seqüência computável x, o que impede a existência de um conjunto maximal;
2. As seqüências típicas de Schnorr são um conjunto mais amplo que as seqüências típicas de Martin-Löf. Isto é conseqüência de uma definição mais "rígida" de conjunto efetivo nulo. Assim, a intersecção dos complementos dos conjuntos efetivos nulos é um conjunto maior que o definido na abordagem de Martin-Löf.

7 Conclusão

Uma importante conseqüência do Teorema 6 é que ele desloca a discussão do conceito de aleatoriedade de uma visão estocástica para uma visão de incompressividade de strings.

Isto tem uma grande importância, em primeiro lugar pela consistência e robustez da definição dada, e em segundo lugar porque esta abordagem, embora baseada em uma função não computável (função parcial recursiva universal), pode ser aproximada pelo uso de um algoritmo compressor universal (assintoticamente ótimo), como por exemplo codificação Lempel-Ziv (programa gzip) ou o algoritmo Burrows-Wheeler (programa bzip2), o que permite algumas medições empíricas, como por exemplo, as propostas em [7, 10, 24].

O conceito de incompressividade, embora originalmente proposto para resolver o problema apresentado neste trabalho, acabou fornecendo uma ferramenta útil em contextos inesperados tais como: teoria de grafos [2]; caracterização de linguagens formais [6, 25]; complexidade computacional [18]; lógica e sistemas formais [9]; raciocínio indutivo e inteligência artificial [31]; avaliação de modelos computacionais, processos e sistemas complexos, particularmente avaliação de modelos de animação gráfica [3, 4, 5, 7]; classificação automática de música (clustering) [10]; biotecnologia e genética [24]; etc.

Muitas aplicações tem resultados bons mesmo que a aproximação computável não seja muito boa, porque os resultados convergem muito rapidamente. Este é o caso da aplicação desta teoria em inteligência artificial, apresentada em [31].

Recentemente, a NASA (agência espacial dos EUA) veiculou artigo pela Internet, em que o pesquisador Frank Corsetti propõem um método para reconhecimento de vida em outros planetas, em imagens enviadas por sondas, baseado na taxa de compressão da imagem, obtida usando-se o programa gzip (veja `http://www.astrobio.net/news/article415.html`). A idéia baseia-se na teoria apresentada neste artigo, ao identificar uma imagem de um ser vivo como algo mais "regular" (mais compressível) que a imagem de uma estrutura mineral.

Estes resultados apresentados aqui tem conseqüências, evidentemente, relacionadas com a existência de geradores de números pseudo-aleatórios (característica normalmente fornecida em linguagens de programação as mais diversas). Em [1] discute-se a existência de algoritmos *seguros* (ou seja, cuja seqüência gerada não possa ser prevista), com desdobramentos importantes na área de criptografia. Em resumo, os autores provam a vinculação entre a existência de tais geradores de números pseudo-aleatórios seguros, e a impossibilidade da computação probabilística melhorar uniformemente a performance das máquinas em relação à computação determinística.

Dois trabalhos recentes (em [3, 7]) apresentam a idéia de compressividade como uma métrica de qualidade de imagem, uma aplicação que é conseqüência da teoria apresentada aqui. Usa-se, neste caso, o compressor bzip2, por exemplo, para obter uma aproximação da complexidade de Kolmogorov que funciona como medida de *similaridade* entre imagens.

Esta idéia de similaridade é apresentada em [24] e usada em duas aplicações. A primeira é o uso desta medida em seqüências de DNA para reconhecimento de genoma mitocon-

drial. Os resultados demonstram que foi possível reconhecer as relações entre três grupos de mamíferos placentários. A segunda aplicação apresentada no artigo preocupa-se em aplicar a medida para a determinação do parentesco de línguas, formando uma hierarquia de linguagens humanas. O experimento baseou-se na tradução da "Declaração Universal dos Direitos do Homem" para 52 línguas diferentes.

Existe um grupo de pesquisadores na área de física, que discute a possibilidade de identificar o próprio Universo como uma máquina de Turing universal [32]. Esta teoria é chamada de *TOE – theories of everything*, e apresenta muitos resultados tomados da área de complexidade de Kolmogorov. Uma das principais discussões nesta teoria é a possibilidade, ou impossibilidade, de prever todos os eventos do Universo, a partir do pressuposto que ele é "simples" em sua origem. Ou seja, o "programa" que foi inserido inicialmente nesta "máquina-universo" era "pequeno".

Embora não seja possível, segundo a teoria da aleatoriedade apresentada neste trabalho, provar que a seqüência de dígitos de um determinado número real representa uma seqüência aleatória, pois a complexidade de Kolmogorov não é computável, é possível definir de forma construtiva um número real que satisfaz esta condição. Trata-se do famoso número Ω, ou *problema da probabilidade de parada*, proposto por Chaitin [8, 9]. Ω é definido como a probabilidade de uma máquina de Turing parar ao executar um programa aleatório (produzido pelo lançamento de uma moeda), $\Omega = \sum_{p \text{ pára}} 2^{-|p|}$. Este número Ω conteria todas as verdades matemáticas construtivas, da forma mais compacta possível, significando que a posse de tal número real permitiria deduzir a validade ou falsidade de todos os enunciados da matemática, inclusive permitiria resolver o famoso *problema da parada* [19].

Para saber mais sobre a teoria e aplicações de complexidade de Kolmogorov consulte [15], que é uma breve introdução à área. Como um estudo complementar aos aspectos apresentados neste trabalho, consulte [6]. Para uma discussão ampla e completa sobre o assunto consulte [26].

Muito se poderia ainda falar sobre aplicações da *complexidade de Kolmogorov*, no entanto, focamos neste trabalho apenas a definição de seqüência aleatória, particularmente a abordagem de Martin-Löf e sua relação com incompressividade de strings binárias.

Duas evidências determinam o mérito da definição de Martin-Löf: a existência de um conjunto efetivo nulo maximal; e a equivalência com a definição via incompressividade. Estas duas propriedades não são satisfeitas pela definição de Schnorr, que representa um conjunto mais amplo de seqüências típicas.

A definição de Martin-Löf e a complexidade de Kolmogorov representam um avanço crucial na solução do problema de definir o conceito de seqüência aleatória. Esta definição resolve todos os problemas que apareceram nas Seções 2 e 3, apresentando-se como uma definição "rígida" de aleatoriedade (ela é uma definição mais "estreita" que a de Wald-Church e a

de Schnorr), fornecendo uma definição consistente com a nossa noção intuitiva de "seqüência aleatória" e, ao mesmo tempo, evitando as dificuldades das outras abordagens, sendo assim matematicamente correta.

8 Agradecimentos

Este trabalho foi parcialmente financiado por CNPq (Projetos HoVer-CAM, GRAPHIT, E-Automaton), FINEP/CNPq (Projeto Hyper-Seed) e CAPES/UFPel.

9 *

Referências

[1] Eric Allender. Some consequences of the existence of pseudorandom generators. *Journal of Computer and System Sciences*, 39:101–124, 1989.

[2] H. Buhrman, Ming Li, J. Tromp, and Paul Vitanyi. Kolmogorov random graphs and the incompressibility method. *SIAM J. Comput.*, 29(2):590–599, 1999.

[3] Carlos A. P. Campani and Paulo Blauth Menezes. On the application of Kolmogorov complexity to the characterization and evaluation of computational models and complex systems. Selecionado para o CISST'04. url: `http://www.ufpel.tche.br/ campani/cisst04.pdf`.

[4] Carlos A. P. Campani and Paulo Blauth Menezes. Characterizing the software development process: A new approach based on Kolmogorov complexity. In Moreno-Díaz, Buchberger, and Freire, editors, *Computer Aided Systems Theory - EUROCAST'2001, 8th International Workshop on Computer Aided Systems Theory*, volume 2178 of *Lecture Notes in Computer Science*, pages 242–256. Springer, 2001.

[5] Carlos A. P. Campani and Paulo Blauth Menezes. Aplicação da complexidade de Kolmogorov na caracterização e avaliação de modelos computacionais e sistemas complexos. In Alfio Martini and David Déharbe, editors, *5th Workshop on Formal Methods*, pages 100–112, Brazil, 2002. Sociedade Brasileira de Computação (SBC), Instituto de Informática/UFRGS.

[6] Carlos A. P. Campani and Paulo Blauth Menezes. Complexidade de Kolmogorov e caracterização da hierarquia de classes de linguagens formais - uma introdução. In Alfio Martini and David Déharbe, editors, *5th Workshop on Formal Methods*, pages 68–83, Brazil, 2002. Sociedade Brasileira de Computação (SBC), Instituto de Informática/UFRGS.

[7] Carlos A. P. Campani and Paulo Blauth Menezes. Evaluating computer animation models with lossy data compression using Kolmogorov complexity. In H. R. Arabnia and Y. Mun, editors, *Proceedings of the 2003 International Conference Imaging Science, Systems, and Technology (CISST'03)*, pages 721–725. CSREA Press, 2003.

[8] Gregory J. Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM*, 22:329–340, 1975.

[9] Gregory J. Chaitin. Godel's theorem and information. *International Journal of Theoretical Physics*, 22:941–954, 1982.

[10] Rudi Cilibrasi, Paul Vitányi, and Ronald de Wolf. Algorithmic clustering of music. Por aparecer. url: `http://www.cwi.nl/ paulv/papers/music.ps`.

[11] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

[12] M. Davis. A note on universal Turing machines. In C. Shannon and J. McCarthy, editors, *Automata Studies*, pages 167–175. Princeton University Press, 1956.

[13] P. Fernandez. *Medida e Integração*. Instituto de Matemática Pura e Aplicada (Impa), Rio de Janeiro, 1996. 202 p.

[14] Peter Gács. Lecture notes on descriptional complexity and randomness. Technical report, Boston University, Computer Science Dept., Boston, Dec 1993.

[15] A. Gammerman and V. Vovk. Kolmogorov complexity: Sources, theory and applications. *The Computer Journal*, 42(4):252–255, 1999.

[16] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes*. Claredon Press - Oxford University Press, Oxford, 2 edition, 1992. 541 p.

[17] Barry James. *Probabilidades: Um Curso em Nível Intermediário*. Instituto de Matemática Pura e Aplicada (Impa), Rio de Janeiro, 1996. 299 p.

[18] T. Jiang, Ming Li, and Paul Vitányi. A lower bound on the average-case complexity of shellsort. *J. Assoc. Comp. Mach.*, 47(5):905–911, 2000.

[19] S. Kleene. *Mathematical Logic*. Wiley, New York, 1967.

[20] Andrei N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag, 1933. Tradução para o inglês de N. Morrison: **Foundations of the Theory of Probability**, Chealsea, 1956.

[21] Andrei N. Kolmogorov. On tables of random numbers. *Sankhya, Series A*, 25:369–376, 1963.

[22] Andrei N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of Information Transmission*, 1:4–7, 1965.

[23] Andrei N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, 14:662–664, 1968.

[24] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul Vitányi. The similarity metric. In *14th ACM-SIAM Symp. Discrete Algorithms (SODA)*, 2003.

[25] Ming Li and Paul Vitányi. A new approach to formal language theory by Kolmogorov complexity. *SIAM J. Comput.*, 24(2):398–410, 1995.

[26] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer, New York, 1997.

[27] Elon Lages Lima. *Espaços Métricos*. Instituto de Matemática Pura e Aplicada (Impa), Rio de Janeiro, 1993. 299 p.

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

[29] Per Martin-Löf. On the concept of a random sequence. *Theory Probability Applyed*, 11:177–179, 1966.

[30] Ernest nagel and James R. Newman. *Prova de Gödel*. Debates. Editora Perspectiva, São Paulo, 1973.

[31] Ray Solomonoff. The discovery of algorithmic probability. *Journal of Computer and System Sciences*, 55(1):73–88, Aug 1997.

[32] Max Tegmark. Does the universe in fact contain almost no information? *Foundation Physics Letters*, 9:25–42, 1996.

[33] V. Uspensky, A. Semenov, and A. Shen. Can an individual sequence of zeros and ones be random? *Russian Math. Surveys*, 45(1):121–189, 1990.

[34] Paul Vitányi. Randomness. In: 'Matematica, Logica, Informatica', Volume 12 do 'Storia del XX Secolo', a ser publicado pelo 'Instituto della Enciclopedia Italiana'. url: `http://www.cwi.nl/ paulv/papers/ency.ps`.

[35] Sérgio B. Volchan. The algorithmic theory of randomness. *The American Mathematical Monthly*, jan 2002.

[36] A. Zvonkin and L. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, 25(6):83–124, 1970.