Finite-state dimensions
Lebesgue measure
1901
Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Algorithmic randomness
Martin-Löf 1966

Lebesgue measure
1901

Computability
Turing, Church 1936

Very Big Picture
Very Big Picture

Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Algorithmic
randomness
Martin-Löf 1966

Computability
Turing, Church 1936

Lebesgue measure
1901

Randomness via
martingales
Schnorr 1971
Very Big Picture

Probability
Kolmogorov 1933
Normal Numbers
Borel 1909
Algorithmic randomness
Martin-Löf 1966
Randomness via martingales
Schnorr 1971
Normality is finite-state randomness
Schnorr & Stimm 1972

Lebesgue measure 1901

Computability
Turing, Church 1936
Very Big Picture

Lebesgue measure 1901

Probability
Kolmogorov 1933

Algorithmic randomness
Martin-Löf 1966

Randomness via martingales
Schnorr 1971

Normal Numbers
Borel 1909

Computability
Turing, Church 1936

Normality is finite-state randomness
Schnorr & Stimm 1972

Hausdorff dimension 1918

Lebesgue measure 1901

Probability
Kolmogorov 1933

Algorithmic randomness
Martin-Löf 1966

Randomness via martingales
Schnorr 1971

Normal Numbers
Borel 1909

Computability
Turing, Church 1936

Normality is finite-state randomness
Schnorr & Stimm 1972

Hausdorff dimension 1918
Very Big Picture

Lebesgue measure 1901

Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Hausdorff
dimension 1918

Algorithmic dimensions
L 2003, AHLM 2007

Algorithmic randomness
Martin-Löf 1966

Computability
Turing, Church 1936

Randomness via martingales
Schnorr 1971

Normality is finite-state randomness
Schnorr & Stimm 1972

Finite-state dimensions
DLLM 2004, AHLM 2007
Normality

\[ \Sigma = \{0, 1, \cdots, b - 1\} \quad (2 \leq b \in \mathbb{N}) \]

For \( S \in \Sigma^\infty \), \( w \in \Sigma^+ \), and \( n \in \mathbb{Z}^+ \),

\[
\text{freq}_n(w, S) = \frac{\left| \left\{ i < n \mid S[i \ldots i + |w| - 1] = w \right\} \right|}{n}
\]

\[ = n \text{-th frequency of } w \text{ in } S \]
Normality

$\Sigma = \{0, 1, \cdots, b - 1\}$ (2 ≤ $b \in \mathbb{N}$)

For $S \in \Sigma^\infty$, $w \in \Sigma^+$, and $n \in \mathbb{Z}^+$,

$$\text{freq}_n(w, S) = \left| \{ i < n \mid S[i \cdots i + |w| - 1] = w \} \right|$$

$$= n\text{-th frequency of } w \text{ in } S$$

**Definition (Borel 1909)**

A sequence $S \in \Sigma^\infty$ is **normal** if

$$(\forall w \in \Sigma^+) \lim_{n \to \infty} \text{freq}_n(w, S) = b^{-|w|}.$$
Normality

\[ \Sigma = \{0, 1, \cdots, b - 1\} \quad (2 \leq b \in \mathbb{N}) \]

For \( S \in \Sigma^\infty \), \( w \in \Sigma^+ \), and \( n \in \mathbb{Z}^+ \),

\[
freq_n(w, S) = \left| \left\{ i < n \mid S[i \ldots i + |w| - 1] = w \right\} \right|_n
= n\text{-th frequency of } w \text{ in } S
\]

**Definition (Borel 1909)**

A sequence \( S \in \Sigma^\infty \) is **normal** if

\[
(\forall w \in \Sigma^+) \lim_{n \to \infty} freq_n(w, S) = b^{-|w|}.
\]

A real number \( \alpha \) is **normal in base** \( b \) if the base-\( b \) expansion of \( \{\alpha\} \) is a normal sequence.
Martin-Löf’s 1966 definition of algorithmic randomness was compelling because it was based on measure theory. Specifically, he defined a sequence \( S \in \{0, 1\}^* \) to be \textit{random} if \( \{S\} \) does not have \textit{algorithmic} \( (\Sigma^0_1) \) Lebesgue measure 0.

Many equivalent characterizations of randomness are now known. All use computability theory. Schnorr’s characterization in terms of martingales is especially useful: Martingales have nice linearity properties. The martingale characterization works at all levels of effectivity: \( \Sigma^0_1 \), computable, \( p \), \( \text{pspace} \), finite-state, etc.
Martin-Löf’s 1966 definition of algorithmic randomness was compelling because it was based on measure theory. Specifically, he defined a sequence $S \in \{0,1\}^*$ to be random if $\{S\}$ does not have algorithmic ($\Sigma^0_1$) Lebesgue measure 0.

Many equivalent characterizations of randomness are now known. All use computability theory.
Martin-Löf’s 1966 definition of algorithmic randomness was compelling because it was based on measure theory. Specifically, he defined a sequence $S \in \{0,1\}^*$ to be random if $\{S\}$ does not have algorithmic ($\Sigma^0_1$) Lebesgue measure 0.

Many equivalent characterizations of randomness are now known. All use computability theory.

Schnorr’s characterization in terms of martingales is especially useful:
Martin-Löf’s 1966 definition of algorithmic randomness was compelling because it was based on measure theory. Specifically, he defined a sequence $S \in \{0, 1\}^*$ to be random if $\{S\}$ does not have algorithmic ($\Sigma^0_1$) Lebesgue measure 0.

Many equivalent characterizations of randomness are now known. All use computability theory.

Schnorr’s characterization in terms of martingales is especially useful:

Martingales have nice linearity properties.

The martingale characterization works at all levels of effectivity: $\Sigma^0_1$, computable, p, pspace, finite-state, etc.
A **martingale** is a function \( d : \{0, 1\}^* \to [0, \infty) \) that satisfies

\[
d(w) = \frac{d(w0) + d(w1)}{2}
\]

for all \( w \in \{0, 1\}^* \).
A **martingale** is a function $d : \{0, 1\}^* \rightarrow [0, \infty)$ that satisfies

$$d(w) = \frac{d(w0) + d(w1)}{2}$$

for all $w \in \{0, 1\}^*$.

A martingale $d$ **succeeds** on a sequence $S \in \{0, 1\}^\omega$ if

$$\limsup_{w \to S} d(w) = \infty.$$
A **martingale** is a function $d : \{0, 1\}^* \rightarrow [0, \infty)$ that satisfies

$$d(w) = \frac{d(w0) + d(w1)}{2}$$

for all $w \in \{0, 1\}^*$.

A martingale $d$ **succeeds** on a sequence $S \in \{0, 1\}^\omega$ if

$$\limsup_{w \to S} d(w) = \infty.$$ 

A martingale $d$ **succeeds strongly** on a sequence $S \in \{0, 1\}^\omega$ if

$$\liminf_{w \to S} d(w) = \infty.$$
A **martingale** is a function \( d : \{0, 1\}^* \rightarrow [0, \infty) \) that satisfies

\[
d(w) = \frac{d(w0) + d(w1)}{2}
\]

for all \( w \in \{0, 1\}^* \).

A martingale \( d \) **succeeds** on a sequence \( S \in \{0, 1\}^\omega \) if

\[
\limsup_{w \to S} d(w) = \infty.
\]

A martingale \( d \) **succeeds strongly** on a sequence \( S \in \{0, 1\}^\omega \) if

\[
\liminf_{w \to S} d(w) = \infty.
\]

\( S^\infty[d] = \text{the success set of } d = \{S \mid d \text{ succeeds on } S\} \).
A **martingale** is a function \( d : \{0, 1\}^* \rightarrow [0, \infty) \) that satisfies

\[
d(w) = \frac{d(w0) + d(w1)}{2}
\]

for all \( w \in \{0, 1\}^* \).

A martingale \( d \) **succeeds** on a sequence \( S \in \{0, 1\}^\omega \) if

\[
\limsup_{w \to S} d(w) = \infty.
\]

A martingale \( d \) **succeeds strongly** on a sequence \( S \in \{0, 1\}^\omega \) if

\[
\liminf_{w \to S} d(w) = \infty.
\]

\( S^\infty[d] \) = the **success set** of \( d = \{S \mid d \text{ succeeds on } S\} \).

\( S_{str}[d]^\infty \) = the **strong success set** of \( d = \{S \mid S \text{ succeeds strongly on } S\} \).
Theorem (Ville 1939)

A set $E \subseteq \{0, 1\}^\omega$ has Lebesgue measure 0 if and only if there exists a martingale $d$ such that $E \subseteq S^\infty[d]$.

Theorem (Schnorr 1971)

A sequence $S \in \{0, 1\}^\omega$ is random in the sense of Martin-Löf (algorithmically random) if and only if there is no lower semicomputable martingale that succeeds on $S$.

Note: $d$ is lower semicomputable if there is a computable function $\hat{d}$: $\{0, 1\}^* \times \mathbb{N} \to \mathbb{Q}$ satisfying the following two conditions for all $w \in \{0, 1\}^*$.

1. For all $t \in \mathbb{N}$, $\hat{d}(x, t) \leq \hat{d}(x, t+1) \leq d(x)$.
2. $\lim_{t \to \infty} \hat{d}(x, t) = d(x)$.
Theorem (Ville 1939)

A set $E \subseteq \{0, 1\}^\omega$ has Lebesgue measure 0 if and only if there exists a martingale $d$ such that $E \subseteq S^\infty[d]$.

Theorem (Schnorr 1971)

A sequence $S \in \{0, 1\}^\omega$ is random in the sense of Martin-Löf (algorithmically random) if and only if there is no lower semicomputable martingale that succeeds on $S$. 

Note: $d$ is lower semicomputable if there is a computable function $\hat{d} : \{0, 1\}^* \times \mathbb{N} \rightarrow \mathbb{Q}$ satisfying the following two conditions for all $w \in \{0, 1\}^*$.

1. For all $t \in \mathbb{N}$, $\hat{d}(x, t) \leq \hat{d}(x, t+1) \leq d(x)$.
2. $\lim_{t \to \infty} \hat{d}(x, t) = d(x)$. 
Theorem (Ville 1939)
A set \( E \subseteq \{0, 1\}^\omega \) has Lebesgue measure 0 if and only if there exists a martingale \( d \) such that \( E \subseteq S^\infty[d] \).

Theorem (Schnorr 1971)
A sequence \( S \in \{0, 1\}^\omega \) is random in the sense of Martin-Löf (algorithmically random) if and only if there is no lower semicomputable martingale that succeeds on \( S \).

Note: \( d \) is lower semicomputable if there is a computable function \( \hat{d} : \{0, 1\}^* \times \mathbb{N} \rightarrow \mathbb{Q} \) satisfying the following two conditions for all \( w \in \{0, 1\}^* \).

1. For all \( t \in \mathbb{N} \), \( \hat{d}(x, t) \leq \hat{d}(x, t + 1) \leq d(x) \).
2. \( \lim_{t \to \infty} \hat{d}(x, t) = d(x) \).
But today we want **finite-state** martingales!

**Example**

![Diagram](image)

\[ \beta(q)(0) = 0.3, \beta(r)(0) = 0.7 \]
\[ \beta(q)(1) = 0.7, \beta(r)(1) = 0.3 \]
But today we want **finite-state** martingales!

Example

![Graph](#)

\[ \beta(q)(0) = .3, \beta(r)(0) = .7 \]
\[ \beta(q)(1) = .7, \beta(r)(1) = .3 \]

\[ d_G(w) = \text{capital that } G \text{ has after } w \text{ if payoffs are fair.} \]
But today we want **finite-state** martingales!

Example

\[
\begin{align*}
\beta(q)(0) &= .3, \beta(r)(0) = .7 \\
\beta(q)(1) &= .7, \beta(r)(1) = .3
\end{align*}
\]

\[d_G(w) = \text{capital that } G \text{ has after } w \text{ if payoffs are fair.}\]

\[d_G(\lambda) = \text{always 1}\]
But today we want finite-state martingales!

Example

\[ \beta(q)(0) = .3, \beta(r)(0) = .7 \]
\[ \beta(q)(1) = .7, \beta(r)(1) = .3 \]

\[ d_G(w) = \text{capital that } G \text{ has after } w \text{ if payoffs are fair.} \]

\[ d_G(\lambda) = 1 \text{ always} \]

\[ d_G(1) = 2(0.7)d_G(\lambda) = 1.4 \]
But today we want finite-state martingales!

Example

\[ \beta(q)(0) = 0.3, \beta(r)(0) = 0.7 \]
\[ \beta(q)(1) = 0.7, \beta(r)(1) = 0.3 \]

\[ d_G(w) = \text{capital that } G \text{ has after } w \text{ if payoffs are fair.} \]

\[ d_G(\lambda) = 1 \text{ always} \]

\[ d_G(1) = 2(0.7)d_G(\lambda) = 1.4 \]

\[ d_G(11) = 2(0.3)d_G(1) = 0.84 \]
But today we want **finite-state** martingales!

Example

\[ \beta(q)(0) = .3, \beta(r)(0) = .7 \]
\[ \beta(q)(1) = .7, \beta(r)(1) = .3 \]

\( d_G(w) \) = capital that \( G \) has after \( w \) if payoffs are fair.

\( d_G(\lambda) = 1 \text{ always} \)

\( d_G(1) = 2(0.7)d_G(\lambda) = 1.4 \)

\( d_G(11) = 2(0.3)d_G(1) = 0.84 \)

\( d_G(110) = 2(0.7)d_G(11) = 1.176 \)
Theorem (Schnorr, Stimm 1972)

A sequence $S \in \Sigma^\omega$ is normal in the sense of Borel (1909) if and only if there is no finite-state gambler that succeeds on $S$. 

\[
\therefore \text{Normality is finite-state randomness!}
\]
Theorem (Schnorr, Stimm 1972)

A sequence $S \in \Sigma^\omega$ is normal in the sense of Borel (1909) if and only if there is no finite-state gambler that succeeds on $S$.

\[ \therefore \text{Normality is finite-state randomness!} \]
They proved even more!

The Schnorr-Stimm dichotomy:

1. If \( S \in \Sigma^\omega \) is not normal, then there exist a finite-state gambler \( G \) and a real \( \alpha > 1 \) such that, for infinitely many prefixes \( w \sqsubseteq S \),

\[
d_G(w) > \alpha^{|w|}.
\]

2. If \( S \in \Sigma^\omega \), then, for every finite-state gambler \( G \) there is a real \( \alpha < 1 \) such that, for every prefix \( w \sqsubseteq S \),

\[
d_G(w) < \alpha^{|\#bets(w)|},
\]

where \( \#bets(w) \) is the number of times \( G \) actually bets on \( w \).
The Schnorr-Stimm dichotomy, more succinctly:

1. If $S$ is not normal then there is a finite-state gambler that makes money at an infinitely-often exponential rate on $S$. 
The Schnorr-Stimm dichotomy, more succinctly:

1. If $S$ is not normal then there is a finite-state gambler that makes money at an infinitely-often exponential rate on $S$.

2. If $S$ is normal, then every finite-state gambler loses money at an exponential rate when it actually bets on $S$. 
Very Big Picture

Lebesgue measure
1901

Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Hausdorff
dimension
1918

Algorithmic dimensions
L 2003, AHLM 2007

Algorithmic
randomness
Martin-Löf 1966

Computability
Turing, Church 1936

Randomness via
martingales
Schnorr 1971

Normality is
finite-state
randomness
Schnorr & Stimm 1972

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Finite-state
dimensions
DLLM 2004, AHLM 2007
Next topic: A very recent quantitative refinement of the Schnorr-Stimm dichotomy (after some background).
Next topic: A very recent quantitative refinement of the Schnorr-Stimm dichotomy (after some background).

Schnorr and Stimm correctly noted that their dichotomy extends routinely to arbitrary probability measures on the alphabet $\Sigma$. 
Next topic: A very recent quantitative refinement of the Schnorr-Stimm dichotomy (after some background).

Schnorr and Stimm correctly noted that their dichotomy extends routinely to arbitrary probability measures on the alphabet $\Sigma$.

A \textbf{(discrete) probability measure} on a nonempty finite set $\Omega$ is a function $\pi : \Omega \to [0, 1]$ satisfying

$$\sum_{w \in \Omega} \pi(w) = 1.$$
Next topic: A very recent quantitative refinement of the Schnorr-Stimm dichotomy (after some background).

Schnorr and Stimm correctly noted that their dichotomy extends routinely to arbitrary probability measures on the alphabet $\Sigma$.

A (discrete) probability measure on a nonempty finite set $\Omega$ is a function $\pi : \Omega \rightarrow [0, 1]$ satisfying

$$\sum_{w \in \Omega} \pi(w) = 1.$$ 

Today $\Omega$ will be $\Sigma^l$ for some $l \in \mathbb{Z}^+$. 
We use the economists’ notation $\Delta(\Omega)$ for the set of all probability measures on $\Omega$. 
We use the economists’ notation $\Delta(\Omega)$ for the set of all probability measures on $\Omega$.

This is a great notation, because $\Delta(\Omega)$ is the set of all points on the $(|\Omega|-1)$-dimensional unit simplex in $|\Omega|$-dimensional Euclidean space.
Let \( w, x \in \Sigma^+ \) be nonempty strings.
Let $w, x \in \Sigma^+$ be nonempty strings.

The **number of block occurrences** of $w$ in $x$ is

$$\# \Box(w, x) = \left| \{ m \leq \frac{|x|}{|w|} - 1 \mid x[m|w|...(m+1)|w| - 1] = w \} \right|.$$ 

Note that $0 \leq \# \Box(w, x) \leq \left\lfloor \frac{|x|}{|w|} \right\rfloor$. 

Jack H. Lutz, 2024
Let \( w, x \in \Sigma^+ \) be nonempty strings.

The **number of block occurrences** of \( w \) in \( x \) is
\[
\#(w, x) = |\{m \leq \frac{|x|}{|w|} - 1 \mid x[m|w|\ldots(m + 1)|w| - 1] = w\}|.
\]
Note that \( 0 \leq \#(w, x) \leq \frac{|x|}{|w|} \).

For \( S \in \Sigma^\omega, n \in \mathbb{Z}^+, \) and \( w \in \Sigma^+ \), the **\( n^{th} \) block frequency** of \( w \) in \( S \) is
\[
\pi_{s,n}(w) = \frac{\#(w, S[0..|w| - 1])}{n}.
\]
Let $w, x \in \Sigma^+$ be nonempty strings.

The **number of block occurrences** of $w$ in $x$ is
$$
\#□(w, x) = |\{m \leq \frac{|x|}{|w|} - 1 \mid x[m|w|...(m + 1)|w| - 1] = w\}|.
$$
Note that $0 \leq \#□(w, x) \leq \frac{|x|}{|w|}$.

For $S \in \Sigma^\omega$, $n \in \mathbb{Z}^+$, and $w \in \Sigma^+$, the $n^{th}$ block frequency of $w$ in $S$ is
$$
\pi_{s,n}(w) = \frac{\#□(w, S[0..n|w| - 1])}{n}.
$$

$$
\pi_{s,n}^{(l)} \overset{\text{def}}{=} \pi_{s,n} \upharpoonright \Sigma^l \in \Delta(\Sigma^l) \text{ is the } n^{th} \text{ empirical probability measure on } \Sigma^l \text{ given by } S.
$$
Let $\alpha \in \Delta(\Sigma), S \in \Sigma^\omega, l \in \mathbb{Z}^+$
Let $\alpha \in \Delta(\Sigma), S \in \Sigma^\omega, l \in \mathbb{Z}^+$

$S$ is $\alpha$-$l$-normal if, for all $w \in \Sigma^l$,

$$\lim_{n \to \infty} \pi_{S,n}(w) = \alpha^{(l)}(w) \overset{\text{def}}{=} \prod_{i=0}^{\lvert w \rvert - 1} \alpha(w[i]).$$
Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, $l \in \mathbb{Z}^+$

$S$ is $\alpha$-$l$-normal if, for all $w \in \Sigma^l$,

$$\lim_{n \to \infty} \pi_{s,n}(w) = \alpha^{(l)}(w) \overset{\text{def}}{=} \prod_{i=0}^{|w|-1} \alpha(w[i]).$$

$S$ is $\alpha$-normal if, for all $l \in \mathbb{Z}^+$, $S$ is $\alpha$-$l$-normal.
Let $\alpha \in \Delta(\Sigma)$, $S \in \Sigma^\omega$, $l \in \mathbb{Z}^+$

S is $\alpha$-$l$-normal if, for all $w \in \Sigma^l$,

$$\lim_{n \to \infty} \pi_{s,n}(w) = \alpha^{(l)}(w) \overset{\text{def}}{=} \prod_{i=0}^{|w|-1} \alpha(w[i]).$$

S is $\alpha$-normal if, for all $l \in \mathbb{Z}^+$, S is $\alpha$-$l$-normal.

S is $l$-normal if S is $\mu$-$l$-normal, where $\mu$ is uniform.
Let $\alpha \in \Delta(\Sigma), S \in \Sigma^\omega, l \in \mathbb{Z}^+$.

$S$ is $\alpha$-$l$-normal if, for all $w \in \Sigma^l$, 

$$\lim_{n \to \infty} \pi_{s,n}(w) = \alpha^{(l)}(w) \overset{\text{def}}{=} \prod_{i=0}^{\lfloor |w| - 1 \rfloor} \alpha(w[i]).$$

$S$ is $\alpha$-normal if, for all $l \in \mathbb{Z}^+$, $S$ is $\alpha$-$l$-normal.

$S$ is $l$-normal if $S$ is $\mu$-$l$-normal, where $\mu$ is uniform.

$S$ is normal if, for all $l \in \mathbb{Z}^+$, $S$ is $l$-normal.
Let $\alpha \in \Delta(\Sigma), S \in \Sigma^\omega, l \in \mathbb{Z}^+$

$S$ is $\alpha$-$l$-normal if, for all $w \in \Sigma^l$, 

$$
\lim_{n \to \infty} \pi_{s,n}(w) = \alpha^{(l)}(w) \overset{\text{def}}{=} \prod_{i=0}^{\lvert w \rvert - 1} \alpha(w[i]).
$$

$S$ is $\alpha$-normal if, for all $l \in \mathbb{Z}^+$, $S$ is $\alpha$-$l$-normal.

$S$ is $l$-normal if $S$ is $\mu$-$l$-normal, where $\mu$ is uniform.

$S$ is normal if, for all $l \in \mathbb{Z}^+$, $S$ is $l$-normal.

**Theorem (Niven and Zuckerman 1951)**

This is equivalent to the “sliding window” definition that we used earlier.
Let $\alpha, \beta \in \Delta(\Omega)$. The **Kullback-Leibler divergence (KL-divergence)** of $\beta$ from $\alpha$ is

$$D(\alpha||\beta) = \mathbb{E}_\alpha \log \frac{\alpha}{\beta}$$
Let $\alpha, \beta \in \Delta(\Omega)$. The **Kullback-Leibler divergence (KL-divergence)** of $\beta$ from $\alpha$ is

$$D(\alpha \parallel \beta) = \mathbb{E}_\alpha \log \frac{\alpha}{\beta}$$

$$= \sum_{w \in \Omega} \alpha(w) \log \frac{\alpha(w)}{\beta(w)},$$

where $\log = \log_2$. 
Properties of KL-Divergence.

1. $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$. 
Properties of KL-Divergence.

1. $D(\alpha \| \beta) \geq 0$, with equality if and only if $\alpha = \beta$.

2. But KL-divergence is not a metric!
Properties of KL-Divergence.

1. $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$.

2. But KL-divergence is not a metric!

3. Entropy is divergence from certainty:

$$H(\alpha) = \mathbb{E}_\alpha D(\pi_w||\alpha).$$
Properties of KL-Divergence.

1. $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$.

2. But KL-divergence is not a metric!

3. Entropy is divergence from certainty:

   $$H(\alpha) = \mathbb{E}_\alpha D(\pi_w || \alpha).$$

4. Mutual information is divergence from independence:

   $$I(\alpha; \beta) = D(\alpha\beta || \gamma),$$

   where $\gamma$ is the joint probability measure of $\alpha$ and $\beta$. 
Properties of KL-Divergence.

1. $D(\alpha||\beta) \geq 0$, with equality if and only if $\alpha = \beta$.

2. But KL-divergence is **not** a metric!

3. Entropy is divergence from certainty:

   $$H(\alpha) = \mathbb{E}_\alpha D(\pi_w||\alpha).$$

4. Mutual information is divergence from independence:

   $$I(\alpha; \beta) = D(\alpha\beta||\gamma),$$

   where $\gamma$ is the joint probability measure of $\alpha$ and $\beta$.

∴ KL-divergence is **fundamental** to Shannon information theory.
Let $l \in \mathbb{Z}^+, S \in \Sigma^\omega, \alpha \in \Delta(\Sigma)$. 

Asymptotic Divergences
Asymptotic Divergences

Let $l \in \mathbb{Z}^+$, $S \in \Sigma^\omega$, $\alpha \in \Delta(\Sigma)$.

The upper $l$-divergence of $\alpha$ from $S$ is

$$Div_l(S||\alpha) = \limsup_{n \to \infty} D(\pi_{s,n}^{(l)}||\alpha^{(l)}).$$
Let \( l \in \mathbb{Z}^+ \), \( S \in \Sigma^\omega \), \( \alpha \in \Delta(\Sigma) \).

The **upper l-divergence** of \( \alpha \) from \( S \) is

\[
\text{Div}_l(S||\alpha) = \limsup_{n \to \infty} D(\pi_{s,n}^{(l)}||\alpha^{(l)}).
\]

The **upper divergence** of \( \alpha \) from \( S \) is

\[
\text{Div}(S||\alpha) = \sup_{l \in \mathbb{Z}^+} \frac{\text{Div}_l(S||\alpha)}{l}.
\]
Theorem (Huang, J. Lutz, Mayordomo, and Stull 2021)

For all $\alpha \in \Delta(\Sigma)$ and $S \in \Sigma^\omega$, the following conditions are equivalent.

1. $S$ is $\alpha$-normal.
2. $\text{Div}(S||\alpha) = 0$. 


A finite-state gambler is a 4-tuple

$$G = (Q, \delta, s, B),$$

where

- $Q$ is a finite set of states,
- $\delta : Q \times \Sigma \to Q$ is the transition function,
- $s \in Q$ is the start state, and
- $B : Q \to \Delta_Q(\Sigma)$ is the betting function.
A **finite-state gambler** is a 4-tuple

\[ G = (Q, \delta, s, B), \]

where

- \( Q \) is a finite set of **states**, 
- \( \delta : Q \times \Sigma \to Q \) is the **transition function** 
- \( s \in Q \) is the **start state**, and 
- \( B : Q \to \Delta_Q(\Sigma) \) is the **betting function**.

Given \( \alpha \in \Delta(\Sigma) \), the \( \alpha \)-payoffs give \( G \) the martingale (capital function)

\[ d_{G,\alpha}(\lambda) = 1, \]

\[ d_{G,\alpha}(wa) = d_{G,\alpha}(w) \frac{B(\delta(w))(a)}{\alpha(a)}. \]
If $\delta(w) = q$ is a state of $G$ in which $B(q) = \alpha$, then $d_{G,\alpha}(wa) = d_{G,\alpha}(w)$ for all $a \in \Sigma$. In this case, we say that $G$ does not bet in state $q$. We thus define the risk that $G$ takes in a state $q$ to be $\text{risk}_{G}(q) = D(\alpha || B(q))$, i.e., the KL-divergence of $B(q)$ from not betting. The total risk that $G$ takes along a string $w \in \Sigma^*$ is $\text{Risk}_{G}(w) = \sum_{u \in \mathcal{R}w} \text{risk}_{G}(\delta(u))$. 
If $\delta(w) = q$ is a state of $G$ in which $B(q) = \alpha$, then $d_{G,\alpha}(wa) = d_{G,\alpha}(w)$ for all $a \in \Sigma$. In this case, we say that $G$ does not bet in state $q$. We thus define the risk that $G$ takes in a state $q$ to be

$$\text{risk}_G(q) = D(\alpha \Vert B(q)),$$

i.e., the KL-divergence of $B(q)$ from not betting.
If $\delta(w) = q$ is a state of $G$ in which $B(q) = \alpha$, then $d_{G,\alpha}(wa) = d_{G,\alpha}(w)$ for all $a \in \Sigma$. In this case, we say that $G$ does not bet in state $q$. We thus define the **risk** that $G$ takes in a state $q$ to be

$$risk_G(q) = D(\alpha \| B(q)),$$

i.e., the KL-divergence of $B(q)$ from not betting. The **total risk** that $G$ takes along a string $w \in \Sigma^*$ is

$$Risk_G(w) = \sum_{u \subseteq w} risk_G(\delta(u))$$
Theorem (Huang, J. Lutz, Mayordomo, and Stull 2021)

1. If $S$ is not $\alpha$-normal, then there is a finite-state gambler $G$ such that, for infinitely many prefixes $w \sqsubseteq S$,

\[ d_{G,\alpha}(w) > 2^{0.99} \text{Div}(S||\alpha)|w|. \]
Theorem (Huang, J. Lutz, Mayordomo, and Stull 2021)

1. If $S$ is not $\alpha$-normal, then there is a finite-state gambler $G$ such that, for infinitely many prefixes $w \sqsubseteq S$,

$$d_{G,\alpha}(w) > 2^{0.99} \text{Div}(S||\alpha)|w|.$$ 

2. If $S$ is $\alpha$-normal, then, for every finite-state gambler $G$, for all but finitely many prefixes $w \sqsubseteq S$,

$$d_{G,\alpha}(w) < 2^{-0.99} \text{Risk}_G(w).$$
Let $G = (Q, \delta, s, B)$ be a finite-state gambler, let $\alpha \in \Delta(\Sigma)$, and let $s \in [0, \infty)$ be a “fairness parameter”. The $s$-$\alpha$-gale of $G$ is the function

$$d^{(s)}_{G,\alpha} : \Sigma^* \to [0, \infty)$$

$$d^{(s)}_{G,\alpha}(w) = \frac{d_G(w)}{|\Sigma|^{|w|} \alpha(w)^s},$$

where $d_G$ is the martingale of $G$. 
For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, let

$$G^\alpha_{FS}(S) = \{s \in [0, \infty) | (\exists \text{ FSG } G)d^{(s)}_{G,\alpha} \text{ succeeds on } S\},$$

$$G^{\alpha,\text{str}}_{FS}(S) = \{s \in [0, \infty) | (\exists \text{ FSG } G)d^{(s)}_{G,\alpha} \text{ succeeds strongly on } S\}.$$
For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, let

$$G_{FS}^\alpha(S) = \{ s \in [0, \infty) \mid (\exists \text{ FSG } G)d_{G,\alpha}^{(s)} \text{ succeeds on } S \},$$

$$G_{FS}^{\alpha,\text{str}}(S) = \{ s \in [0, \infty) \mid (\exists \text{ FSG } G)d_{G,\alpha}^{(s)} \text{ succeeds strongly on } S \}.$$

**Definition (Dai, Lathrop, J. Lutz, and Mayordomo 2004)**

The **finite-state $\alpha$-dimension** of $S$ is

$$\dim_{FS}^\alpha(S) = \inf G_{FS}^\alpha(S).$$
For each $\alpha \in \Delta(\Sigma)$ and each $S \in \Sigma^\omega$, let

$$G^\alpha_{FS}(S) = \{s \in [0, \infty) \mid (\exists \text{ FSG } G)d_{G,\alpha}^{(s)} \text{ succeeds on } S\},$$

$$G^{\alpha,\text{str}}_{FS}(S) = \{s \in [0, \infty) \mid (\exists \text{ FSG } G)d_{G,\alpha}^{(s)} \text{ succeeds strongly on } S\}.$$

Definition (Dai, Lathrop, J. Lutz, and Mayordomo 2004)

The **finite-state $\alpha$-dimension** of $S$ is

$$\dim_{FS}^\alpha(S) = \inf G^\alpha_{FS}(S).$$

Definition (Athreya, Hitchcock, J. Lutz, and Mayordomo 2007)

The **strong finite-state $\alpha$-dimension** of $S$ is

$$\Dim_{FS}^\alpha(S) = \inf G^{\alpha,\text{str}}_{FS}(S).$$
Observations.

1. If $G$ never bets in any state and $s > 1$, then $d_{G,\alpha}^{(s)}(w) = |\Sigma|^{(s-1)|w|}$ for all $w \in \Sigma^*$. This implies that, for all $S \in \Sigma^\infty$, $0 \leq \dim_{FS}^\alpha(S) \leq \Dim_{FS}^\alpha(S) \leq 1$.

2. Since $d_{G,\alpha}^{(1)}(w) = d_{G,\alpha}(w)$, every $\alpha$-normal sequence $S \in \Sigma^\omega$ has $\dim_{FS}^\alpha(S) = 1$.

3. (Bourke, Hitchcock, and Vinodchandran 2005) The Schnorr-Stimm dichotomy tells us that, if $S$ is not $\alpha$-normal, then $\dim_{FS}^\alpha(S) < 1$. 

Jack H. Lutz, 2024
Observations.

1. If $G$ never bets in any state and $s > 1$, then $d_{G,\alpha}^{(s)}(w) = |\Sigma|^{(s-1)|w|}$ for all $w \in \Sigma^*$. This implies that, for all $S \in \Sigma^\infty$,

$$0 \leq \dim_{FS}^\alpha(S) \leq \Dim_{FS}^\alpha(S) \leq 1.$$ 

2. Since $d_{G,\alpha}^{(1)} = d_{G,\alpha}$, every $\alpha$-normal sequence $S \in \Sigma^\omega$ has $\dim_{FS}^\alpha(S) = 1$. 
Observations.

1. If $G$ never bets in any state and $s > 1$, then $d_{G,\alpha}^{(s)}(w) = |\Sigma|^{(s-1)|w|}$ for all $w \in \Sigma^*$. This implies that, for all $S \in \Sigma^\infty$,

$$0 \leq \dim_{FS}^\alpha(S) \leq \dim_{FS}^\alpha(S) \leq 1.$$ 

2. Since $d_{G,\alpha}^{(1)} = d_{G,\alpha}$, every $\alpha$-normal sequence $S \in \Sigma^\omega$ has $\dim_{FS}^\alpha(S) = 1$.

3. (Bourke, Hitchcock, and Vinodchandran 2005)

The Schnorr-Stimm dichotomy tells us that, if $S$ is not $\alpha$-normal, then $\dim_{FS}^\alpha(S) < 1$. 
Hence a sequence $S \in \Sigma^\omega$ is $\alpha$-normal if and only if $\dim_{FS}^\alpha(S) = 1$. 
Hence a sequence $S \in \Sigma^\omega$ is $\alpha$-normal if and only if $\dim^\alpha_{FS}(S) = 1$.

The finite-state world is the only world that we know of where dimension 1 implies randomness.
Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Hausdorff
dimension
1918

Algorithmic dimensions
L 2003, AHLM 2007

Algorithmic
computability
Turing, Church 1936

Algorithmic
dimension
1901

Lebesgue measure

Lebesgue measure
1901

Algorithmic
computability
Turing, Church 1936

Algorithmic
dimension
1901

Hausdorff
dimension
1918

Algorithmic dimensions
L 2003, AHLM 2007

Algorithmic randomness
Martin-Löf 1966

Randomness via
martingales
Schnorr 1971

Randomness via
martingales
Schnorr 1971

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Finite-state
dimensions
DLLM 2004, AHLM 2007

Finite-state
dimensions
DLLM 2004, AHLM 2007

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Randomness via
martingales
Schnorr 1971

Randomness via
martingales
Schnorr 1971

Probability
Kolmogorov 1933

Normal Numbers
Borel 1909

Hausdorff
dimension
1918

Algorithmic dimensions
L 2003, AHLM 2007

Algorithmic randomness
Martin-Löf 1966

Randomness via
martingales
Schnorr 1971

Normality is
finite-state
dimension 1
SS 1972, BHV 2005

Finite-state
dimensions
DLLM 2004, AHLM 2007

Jack H. Lutz, 2024
Thank you!

Thanks to CBMS and to NSF grants 2329555 and 1900716.