Normal Numbers

A Short Proof of the Borel Normal Number Theorem

Blair Drummond

July 12, 2017

University of Ottawa

What's a Normal Number?

What's a Normal Number?

A number $x \in [0, 1]$ is said to be *Normal* with respect to base-b if it can be written in an infinite base-b expansion where the asymptotic frequency of all terms are equal.

Note that, in this interpreation,

$$0.5 = 0.5000000000 \cdots = (5, 0, 0, 0, \dots)$$

We will be describing these expansions as infinite words over the alphabet $\{0, \dots, b-1\}$.

What's a Normal Number?

Some notation

- Denote by Λ the alphabet $\{0, \dots, b-1\}$
- Denote by Λ^* the finite length words over Λ
- Denote by Λ^∞ the infinite length words over Λ
- Let $\delta_i : \Lambda \to \{0, 1\}$ be defined to be

$$\delta_i(j) = \begin{cases} 1 & i = j \\ 0 & otherwise \end{cases}$$

What's a Normal Number?

Conventions

Throughout, we will have some conventions

- ω_i is a letter in Λ
- ω' is a finite length word
- $\omega = (\omega_1, \omega_2, \dots)$ is an infinite word

Λ^* is a monoid under concatenation, and we can act on Λ^∞ by concatenating finite words on the left.

What's a Normal Number?

In this language, x is normal with respect to base-b, iff

$$x = \sum_{n \in \mathbb{N}} \omega_n b^{-(n+1)}$$

and if, asymptotically, the occurrences of all $\omega_i \in \Lambda$ are of equal frequency. That is to say,

$$\forall i \in \Lambda. \ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_i(\omega_k) = \frac{1}{b}$$

This base-b expansion is unique unless x is b-rational, in which case we aren't interested in it anyway.

What's a Normal Number?

We say that $x \in [0, 1]$ is *Normal* if it is normal with respect to **every** base b.[1]

We can extend this to $\mathbb{R}$, as x is normal iff the fractional component is normal[2].

[1] Sometimes this is also called "absolute normality"

[2] The integer portion of a real number has a finite expansion in base b, and hence doesn't affect the limiting average.

Borel Normal Number Theorem

Borel Normal Number Theorem

Borel's Normal Number Theorem

Almost all real numbers are normal.

Almost all(?)

This concept comes from Measure Theory, and it meant to capture the notion of the distribution of “mass” in a space.

A measure space is simply a triple $(\Omega, \mathcal{F}, \mu)$, where

- Ω is our state-space
- $\mathcal{F} \subset \mathcal{P}(\Omega)$ contains our measurable sets[3]
- $\mu : \mathcal{F} \to [0, \infty]$ is a map which assigns mass to measurable sets.

[3] In general, not every set can be measured in an appropriate way.

Borel Normal Number Theorem

Some important conditions are imposed on $\mathcal{F}$ and μ. First, $\mathcal{F}$ is a σ-algebra[4], meaning that

- $\mathcal{F}$ contains $\emptyset$ and Ω
- $\mathcal{F}$ is closed under complements
- $\mathcal{F}$ is closed under **countable** unions and intersections

And μ respects these operations in that μ is countably addative, meaning that for a family of pairwise disjoint sets $\{A_i\}_{i \in \mathbb{N}}$,

$$\mu(\emptyset) = 0 \quad \text{and} \quad \mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mu(A_i)$$

[4] Often we require that it satisfy an extra condition about subsets of a set of measure 0, but we aren't going to worry about that.

Lebesgue Measure

This is the measure that captures the idea of length, area, or volume. On $\mathbb{R}$, it is length. We can also restrict the measure to $[0, 1]$ instead of all of $\mathbb{R}$.

Probability Measures

Every measure μ where $\mu(\Omega) = 1$ is a probability measure. Usually probability measures are denoted by P rather than μ.

Borel Normal Number Theorem

In this context, the statement "Almost all real numbers are normal", means that, with respect to the Lebesgue measure[5],

$$\mu(\{\ x \in \mathbb{R}\ :\ x \text{ is not normal }\}) = 0$$

We are going to restrict our attention to the interval $[0, 1]$ (so that we get a probability measure), and then we can also say

$$P(\{\ x \in [0,1]\ :\ x \text{ is normal }\}) = P(x \in [0,1] \text{ is normal}) = 1$$

[5]We haven't established that this set is measurable, but this will resolve itself

Ergodicity

Ergodic Theory

To show this, we are going to leverage a powerful result: *Birkhoff's Ergodic Theorem*.[6]

Before we go into that, we have to say what an ergodic transformation is. [2]

[6] I found this proof of the Borel Normal Number theorem in Billingsley's book "Ergodic Theory and Information" [1]

Ergodicity

Measure Preserving Transformation

A map[7] $T : \Omega \to \Omega$ is called Measure Preserving if

$$\forall A \in \mathcal{F}.\ P(T^{-1}(A)) = P(A)$$

Note that T needn't be invertible for this to be true; T^{-1} is just a pre-image.

T Invariance

A set is called T invariant if

$$P(T^{-1}A\Delta A) = 0, \quad \text{where} \quad A\Delta B = (A \setminus B) \cup (B \setminus A)$$

[7] Actually, we can only talk about measurable maps. We need $T^{-1}A \in \mathcal{F}$

Ergodicity

A map $T : \Omega \to \Omega$ is called *ergodic* with respect to a probability measure if T is measure preserving and if the only T-invariant sets have probability 1 or probability 0. [8]

This condition amounts to saying that T doesn't fix any non-trivial subsets of Ω. Now, the remarkable result:

[8] If you aren't in a probability space, you instead talk about sets of full measure or measure 0.

Ergodicity

Birkhoff Ergodic Theorem

If T is ergodic and $\mathbb{E}[f]$ exists, then for almost all ω,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(\omega)) = \mathbb{E}[f]$$

I.e. You can compute the expected value of f by taking a sample average along a randomly chosen orbit of T.

The Proof

The Proof

The Punchline

Now, recall that $\omega = (\omega_0, \omega_1, \dots)$ is normal base b if

$$\forall i \in \Lambda. \ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_i(\omega_k) = \frac{1}{b}$$

Seem familiar?

The Proof

Denote by T the shift, $(\omega_0, \omega_1, \ldots) \mapsto (\omega_1, \omega_2, \ldots)$, and notice that[9]

$$\frac{1}{n}\sum_{k=0}^{n-1} \delta_i(\omega_k) = \frac{1}{n}\sum_{k=0}^{n-1} \delta_i(T^k(\omega))$$

And

$$\mathbb{E}[\delta_i] = P(\omega_j = i) = \frac{1}{b}$$

[9] Abusing our definition of δ_i slightly

The Proof

So, if we can show that T is ergodic, then we can show that for all *fixed* bases b, for all i, for almost all ω,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_i(\omega_k) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_i(T^k(\omega)) = \mathbb{E}[\delta_i] = \frac{1}{b}$$

Which shows that almost every number is normal with respect to base b.

Potential Problem?

What that shows is that almost all numbers are normal with respect to each base b separately, but are almost all numbers normal with respect to all of them simultaneously?

We're OK

Denote by S_b the numbers normal with respect to base b. Then x is normal if $x \in \bigcap_{b>1} S_b$. But we can show that the countable intersection of full measure sets is again of full measure.

$$P\left(\bigcap_{b>1} S_b\right) = 1 - P\left(\bigcup_{b>1} S_b^c\right)$$

The Proof

We see that

$$0 \leq P\left(\bigcup_{b>1} S_b^c\right) \leq \sum_{b>1} P\left(S_b^c\right) = \sum_{b>1} 0 = 0$$

So

$$P\left(\bigcap_{b>1} S_b\right) = 1 - P\left(\bigcup_{b>1} S_b^c\right) = 1$$

So, if T is ergodic, then for every fixed b, all numbers are normal with respect to b, and by this result, this makes them normal, which will prove our result.

Ergodicity of the Shift

Proof that T is Ergodic

We're actually going to show something stronger than Ergodicity.
We prove this in steps.

- Show that T is measure preserving on Λ^*
- Conclude that T is measure preserving on Λ^∞
- Show that T is mixing on Λ^*
- Conclude that T is mixing on Λ^∞
- Deduce that T is ergodic (mixing implies ergodic)

Ergodicity of the Shift

Mixing

We're going to show that T is (strongly) mixing, meaning that for all A and B

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B)$$

If T is mixing, then noting that for invariant sets A, $T^{-1}A = A$ on a set of full measure (on both sides). Setting $B = A$ for such an invariant set, we get that $P(A) = P(A)^2$, so $P(A) = 0$ or 1. So **mixing** $\Rightarrow$ **ergodic**.

Ergodicity of the Shift

Why show that it's mixing?

Because we want to leverage two theorem's from Billingsley's book "Ergodic Theory and Information"[1]. These theorems allow us to show that T is measure preserving and mixing by looking at the behavior of T on words with a finite amount of information.

"finite amount of information"?

One last bit of notation:

Ergodicity of the Shift

Cylinder Sets

Denote by $\Gamma \subset \mathcal{P}(\Lambda^\infty)$ which fixes finitely many characters ω_i, and lets the others vary. For example, a set in Γ might look like

$$(* \ldots \omega' * \ldots)$$

$$:=$$

$$(*, *, *, \ldots, *, \omega_n, \omega_{n+1}, \ldots, \omega_{n+k}, *, \ldots)$$

$$:=$$

$$\prod_{i=0}^{n-1} \Lambda \times \prod_{i=0}^{k} \{\omega_{n+i}\} \times \prod_{i \in \mathbb{N}} \Lambda$$

These are often called *cylinder sets*.

Ergodicity of the Shift

The cylinder sets generate the measurable sets of Λ^∞, and the behavior of P on these sets determines behavior on all of the σ-algebra:

Billingsley: Thereom 1.1

If $P(T^{-1}A) = P(A)$ for every cylinder set A, then T is measure preserving on every measurable set.

Billingsley: Thereom 1.2

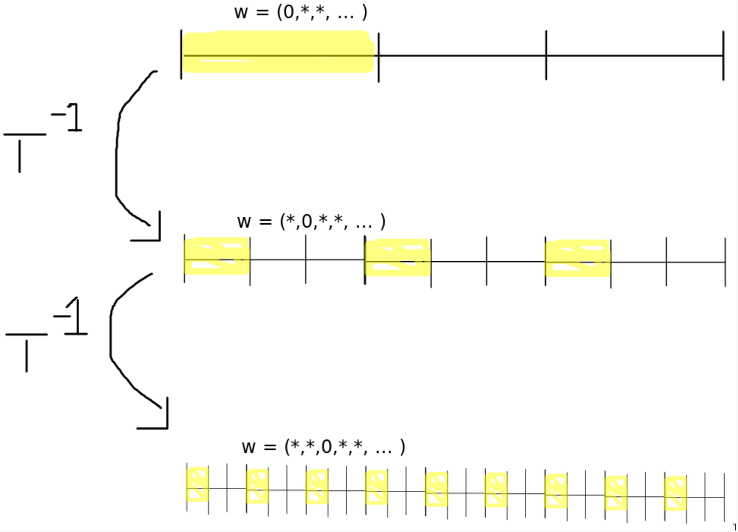
If for all cylinder sets A and B,

$$\lim_{n \to \infty} P(A \cap T^{-n}B) = P(A)P(B)$$

then T is mixing.

Ergodicity of the Shift

Showing that T is measure preserving is easy. Intuitively:



Ergodicity of the Shift

$P((*,\dots,*,\omega_n,*,\dots))$ is just $\sum_{i=1}^{b^n}\frac{1}{b^{n+1}}=\frac{1}{b}$ this formula is invariant when you alter n, and

$$\begin{aligned}
P((*\dots\omega_n,\omega_{n+1},*\dots)) &= P((*\dots\omega_n,*,\dots)\cap(*\cdots*,\omega_{n+1}*\dots)) \\
&= P((*\dots\omega_n,*\dots))P((*\cdots*,\omega_{n+1}*\dots)) \\
&= \frac{1}{b^2}
\end{aligned}$$

The same is true for all cylinders with n fixed entries: they have probability $\frac{1}{b^n}$.

Ergodicity of the Shift

T is measure preserving

T^{-1} simply acts on the cylinder sets by adding a $*$ to the front, which, as described, doesn't impact the associated probability. So T is measure preserving on the cylinder sets, so we conclude from Theorem 1.1 that it is measure preserving.

Next, we show that it is (strongly) mixing on the cylinder sets.

Ergodicity of the Shift

Every cylinder has only finitely many fixed values, so, when looking at

$$\lim_{n\to\infty} P(A \cap T^{-n}B) = P(A)P(B)$$

we can see that past the last fixed element of the cylinder A, $T^{-k}B$ and A become independent.

A	$(*, *, \ldots, \omega_n, \ldots, \omega_{n+k}, *, *, \ldots)$
B	$(*, *, \ldots, \lambda_m, \ldots, \lambda_{m+l}, *, *, \ldots)$
$T^{-n-k-1}B$	$(* \ldots \lambda_{m+n+k+1} \ldots \lambda_{m+n+k+l+1} * \ldots)$
$A \cap T^{-n-k-1}B$	$(* \cdots * \omega' * \cdots * \lambda' * \ldots)$

Ergodicity of the Shift

As you take more preimages of $T^{-n-k-1}B$, the λ' columns just keep moving right -away from ω'-. By the same argument as we made earlier, the probability associated to this term is $\frac{1}{b}$ to the power of $\mathit{length}(\omega') + \mathit{length}(\lambda')$, which is exactly $P(A)P(B)$.

So, T is mixing on the cylinder sets. Applying Billingsley's Theorem 1.2, we conclude that T is mixing, and deduce that T is ergodic. From the argument we laid out, it follows that almost every number in $[0, 1]$ is normal, and hence almost every real is normal. $\blacksquare$

References

[1] Patrick Billingsley. "Ergodic theory and information". In: (1965).

[2] Michael Brin and Garrett Stuck. *Introduction to dynamical systems*. Cambridge university press, 2002.

[3] Joseph L Doob. "The development of rigor in mathematical probability (1900-1950)". In: *The American Mathematical Monthly* 103.7 (1996), pp. 586–595.

[4] Jeffrey S Rosenthal. *A first look at rigorous probability theory*. World Scientific Publishing Co Inc, 2006.

Questions?