“What Does a Typical Normal Number Look Like?,” and Other Enchanting Tales

Davar Khoshnevisan
University of Utah
davar@math.utah.edu
http://www.math.utah.edu/~davar

Our High-Entropy Programme:
(Time permitting)

Normal numbers
$\Rightarrow$ uniform sampling
$\Rightarrow$ uniform sampling from non-normals
$\Rightarrow$ entropy/dimension for non-normals

Why uniform sampling?
Laplace's maximum-entropy principle

Normal Numbers

Choose a number $x \in [0, 1]$ and write it out, in decimal form, as

$$x = 0.x_1 x_2 x_3 \cdots ,$$

where the x_j's are integers between 0 and 9. E.g., $x = 0.5302$. The number x is a "normal number" (aka "simply-normal number") if the asymptotic fraction of every digit in its expansion is $\frac{1}{10}$. [Not to be mistaken with the "normal distribution."]

Some Questions

1. Can you construct a normal number?
   (Doable but requires thought)

2. Is there an algorithm for deciding when a given number is normal?
   (Open for about 100 years)

To see why the latter is a tough problem, consider the following surprising fact:

Theorem 1 (D. G. Champernowne, 1933) *The following is a normal number:*

$$x = 0.123456789101112131415161718 \cdots .$$

Several proofs exist, but none are overtly simple. Can you at least find an intuitive explanation?

The existing literature contains some sufficient conditions for normality, but as far as I know it is not known whether any of the following is normal: $\pi/4$, $e/3, \cdots$.

One might be tempted to think that normal numbers are rare. Quite the opposite is true, though.

Theorem 2 (É. Borel, 1904) *The set of non-normals has zero length.*

To understand why, suppose X is a random variable (r.v.) that is selected *uniformly at random* from $[0, 1]$ ($\mathcal{U}([0, 1])$). That is, for "all" subsets A of $[0, 1]$,

$$\Pr\{X \in A\} = \text{Length}(A).$$

Therefore, Borel's theorem in fact says:

Theorem 3 *If X is $\mathcal{U}([0, 1])$, then with probability one X is normal.*

So, given an honest random-number generator, we could construct all manners of normal numbers. This can be turned around to give you a fitness test for your random number generator! (Statistics+Cramér's theorem of large deviations)

Lemma 4 *If* $X = 0.X_1, X_2 \ldots$ *is* $\mathcal{U}([0,1])$*, then the* X_j*'s are independent, each taking values* $0, \ldots, 9$ *with prob.* $\frac{1}{10}$ *each.*

Proof. Binary (actually 10-ary) search. $\square$

Borel's theorem follows from this and the law of large numbers (A. N. Kolmogorov, 1933): Let $Y_j = 1$ if $X_j = 0$; else, $Y_j = 0$. Then, the Y_j's are independent rv's, and $\mathsf{E}[Y_j] = \frac{1}{10}$. By the law of large numbers, with probability one,

$$\lim_{N \to \infty} \underbrace{\frac{Y_1 + \cdots + Y_N}{N}}_{\text{fraction of } 0\text{'s}} = \frac{1}{10}.$$

Ditto for the (asymptotic) fraction of 1's, 2's, $\ldots, 9$'s.

Non-Normal Numbers

Here is a "nice" class of non-normal numbers: Suppose $\vec{p} = (p_0, \ldots, p_9)$ is a probability vector; i.e., $p_j \in [0, 1]$ and $p_0 + \cdots + p_9 = 1$. Define $N(\vec{p})$ to be the collection of all points $x \in [0, 1]$ such that the asymptotic fraction of j is p_j ($j = 0, \ldots, 9$).

If $p_0 = \cdots = p_9$, then $N(\vec{p})$ is the collection of all normal numbers, and we have seen that in that case $N(\vec{p})$ has full length; i.e., its complement has zero length.

For all other probability vectors $\vec{p}$, $N(\vec{p})$ has zero length. Nevertheless, we can still draw "uniformly" at random from $N(\vec{p})$. Here is how:

Suppose $X_1, X_2, \ldots$ are independent r.v.'s taking the values $0, \ldots, 9$ with probabilities $p_0, \ldots, p_9$, respectively. Now define X to be the random number in $[0, 1]$ whose ith decimal point is X_i; i.e.,

$$X = X_1 \cdot 10^{-1} + X_2 \cdot 10^{-2} + \cdots .$$

What does the distribution of the r.v. X look like? For one, an appeal to the law of large numbers shows that

$$\Pr\{X \in N(\vec{p})\} = 1.$$

So we have described a way to sample randomly from $N(\vec{p})$. But why is it "uniform"?

Choose and fix any point $z \in N(\vec{p})$, and write z, in decimal form, as $z = 0.z_1 z_2 \ldots$. By independence,

$$\begin{aligned}
&\Pr\{X_1 = z_1, \ldots, X_n = z_n\} \\
&= \Pr\{X_1 = z_1\} \times \cdots \times \Pr\{X_n = z_n\} \\
&= p_0^{f_n(0)} \times \cdots \times p_9^{f_n(9)},
\end{aligned}$$

where $f_n(i)$ is the number (*$\underline{\underline{f}}$requency*) of times that $z_1, \ldots, z_n$ equal to i. Now take logs:

$$\begin{aligned}
&\log \Pr\{X_1 = z_1, \ldots, X_n = z_n\} \\
&= f_n(0) \log p_0 + \cdots + f_n(9) \log p_9.
\end{aligned} \tag{1}$$

Because $z \in N(\vec{p})$, the asymptotic fraction of i in the expansion of z is p_i; i.e.,

$$\lim_{n \to \infty} \frac{f_n(i)}{n} = p_i, \qquad \text{for all } i = 0, \ldots, 9.$$

Plug in (1) find:

$$\lim_{n\to\infty} \frac{1}{n} \log \Pr\{X_1 = z_1, \ldots, X_n = z_n\} \\ = p_0 \log p_0 + \cdots + p_9 \log p_9. \tag{2}$$

The thermodynamic *entropy* of the probability vector $\vec{p}$ is simply the absolute-value of the right-hand side; i.e.,

$$\mathrm{Ent}(\vec{p}) = -p_0 \log p_0 - \ldots - p_9 \log p_9.$$

($p_i \leq 1 \;\therefore \log p_i \leq 0 \;\therefore \mathrm{Ent}(\vec{p}) \geq 0$.)

Thus, we can think of (2) as

$$\Pr\{X_1 = z_1, \ldots, X_n = z_n\} \approx 10^{-n\mathrm{Ent}(\vec{p})}.$$

But the left-hand side is just about the same as the probability that X is within 10^{-n} of z. This "shows" that for any $z \in N(\vec{p})$ fixed, and all large n,

$$\Pr\left\{|X - z| \leq 10^{-n}\right\} \approx 10^{-n\mathrm{Ent}(\vec{p})}.$$

"Thus," for all sufficiently small $\varepsilon > 0$,

$$\mathsf{Pr}\left\{|X - z| \le \varepsilon\right\} \approx \varepsilon^{\mathsf{Ent}(\vec{p})}. \tag{3}$$

Because the right-hand side does not depend on z, this justifies (somewhat) the notion that X is uniformly distributed on $N(\vec{p})$ (why? More importantly, why only somewhat?).

Although my "derivation" of (3) has some logical holes in it, these holes can be patched up; (3) itself is entirely true if you interpret "$\approx$" appropriately.

Next is a happy consequence of (3) [in case you have heard of the terms to follow]:

Theorem 5 (H. G. Eggleston, 1949) *For any probability vector* $\vec{p} = (p_0, \ldots, p_9)$,

$$\mathrm{dim}_{\mathcal{H}}\, N(\vec{p}) = \mathrm{Ent}(\vec{p}).$$

Here $\mathrm{dim}_{\mathcal{H}}\, F$ stands for the Hausdorff–Besicovitch (often called fractal) dimension of a set F. P.S. The same formula is valid for the other fractal dimension ("packing") too.

$\mathcal{U}([0,1])$ via Entropy

Why does choosing uniformly work in some instances?

I close by introducing another connection between $\mathcal{U}([0,1])$ and entropy. This connection was originated by P.-S. Laplace (1810's), and is called the "maximum entropy law," as well as the "method of maximum probabilities."

First, some undergraduate probability:

If $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, then f is a so-called “probability density function” or pdf. A random variable X has pdf f if for “all” A,

$$\Pr\{X \in A\} = \int_A f(x)\,dx.$$

If X is $\mathcal{U}([0,1])$, then its pdf is

$$f_{\text{unif}}(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

If f is a pdf, then its entropy is

$$\mathsf{Ent}(f) = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx,$$

where $0 \cdot \ln 0 := 0$. (This is a continuous version of the entropy we saw earlier.) Thus, for example,

$$\mathsf{Ent}\,(f_{\mathsf{unif}}) = 0.$$

The (informal) “law of maximum entropy” states that if you wish to predict the pdf of a r.v. X, then you should maximize entropy. If there is further info, then take that into account while finding the max.

Now suppose we know that we have ourselves an unknown pdf f on $[0,1]$. What is a good guess for f? Because we know only that f is a pdf on $[0,1]$, the "most sensible guess" is f_{unif}.

The maximum-entropy law confirms this: Note that $h(x) = 1 - x + x \ln x$ ($x \geq 0$) is minimized at $x = 1$ with $h(1) = 0$. I.e.,

$$-x \ln x \leq 1 - x, \qquad \text{for all } x \geq 0.$$

Let $x := f(t)$ to deduce that for any pdf f on $[0,1]$,

$$-f(t) \ln f(t) \leq 1 - f(t), \qquad \text{for all } t \in [0,1].$$

Integrate this over all $t \in [0,1]$ now. Because $\int_0^1 f(t)\,dt = 1$, this shows that $\text{Ent}(f) \leq 0 = \text{Ent}(f_{\text{unif}})$!