A Note About Singular Distributions

by Jeffrey S. Rosenthal, www.probability.ca

Let $Z_1, Z_2, \ldots$ be i.i.d. equal to 1 with probability 2/3, or 0 with probability 1/3.

Let $Y = \sum_{n=1}^{\infty} Z_n 2^{-n}$, so that the base-2 expansion of Y is $0.Z_1 Z_2 \ldots$ (base 2).

What kind of random variable is Y?

Well, clearly $\mathbf{P}(Y = c) = 0$ for each fixed real number c, since the infinitely-long base-2 expansion of Y would all have to be "exactly right" to equal c, which has probability 0. So, Y has no discrete component.

But could there be a density function f with $\mathbf{P}(a < Y < b) = \int_a^b f(x)\,dx$ for all $a < b$? Actually, no, there couldn't! Why not?

Well, for any real number r, let

$$\mathbf{FR}(r) \;=\; \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} d_i(r)$$

where $d_i(r)$ is the i^{th} digit in the base-2 expansion of r. That is, $\mathbf{FR}(r)$ is the limiting fraction of 1's in the base-2 expantion of r.

(Strictly speaking, some special real numbers have <u>two</u> different possible base-2 expansions, e.g. $3/4 = 0.1100000\ldots = 0.1011111\ldots$. In such cases, we could specify that we always choose the <u>non-terminating</u> option. Anyway this is not too important.)

First consider $U \sim \text{Uniform}[0,1]$. The base-2 expansion of U has each digit <u>equally</u> likely to be 0 or 1 (since e.g. U is equally likely to be in $[0, 1/2)$ or $[1/2, 1]$, and then equally like to be in $[0, 1/4)$ or $[1/4, 1/2)$, etc.). So, if $U \sim \text{Uniform}[0,1]$, then $\mathbf{P}(\mathbf{FR}(U) = 1/2) = 1$. That is, for U, the limiting fraction of 1's is always 1/2.

Next, let $S_{2/3} = \{x \in [0,1]; \mathbf{FR}(x) = 2/3\}$ be the collection of numbers with limiting fraction of 1's equal to 2/3.

Then, by the above, $\mathbf{P}(U \in S_{2/3}) = 0$.

However, by the construction of Y above, we must have $\mathbf{P}(Y \in S_{2/3}) = 1 > 0$.

Now, <u>if</u> Y had a density function f, then by definition, for all $a < b$,

$$\mathbf{P}(a \le Y \le b) = \int_a^b f(x)\,dx\,,$$

i.e.

$$\mathbf{P}(Y \in [a,b]) = \int_{[a,b]} f(x)\,dx\,.$$

It would then follow that for any (measurable) subset A,

$$\mathbf{P}(Y \in A) = \int_A f(x)\,dx\,.$$

In particular, it would be true that

$$\mathbf{P}(Y \in S_{2/3}) = \int_{S_{2/3}} f(x)\, dx\,. \tag{$*$}$$

But here's the problem. These integrals are all with respect to "dx", which corresponds to the uniform distribution (formally: Lebesgue measure). And, integrating any function f over any set with zero uniform probability is always zero. For example, if $A = \{1/2\}$ and $B = \{1/3, 3/4\}$, then $\mathbf{P}(U \in A) = 0$, and $\mathbf{P}(U \in B) = 0$, so

$$\int_A f(x)\, dx = \int_{1/2}^{1/2} f(x)\, dx = 0\,; \quad \text{and} \quad \int_B f(x)\, dx = \int_{1/3}^{1/3} f(x)\, dx + \int_{3/4}^{3/4} f(x)\, dx = 0\,.$$

In this case, since $\mathbf{P}(U \in S_{2/3}) = 0$, we must have

$$\int_{S_{2/3}} f(x)\, dx = 0\,.$$

But we know that $\mathbf{P}(Y \in S_{2/3}) = 1 > 0$. This gives a contradiction to $(*)$. Hence, Y does <u>not</u> have any density function f with respect to the usual "dx" (i.e. Lebesgue measure).

Distributions like this one for Y, which do not have any discrete component but also do not have any density function component, are called <u>singular</u> distributions. They are a whole different kind of distribution, never even imagined in most undergraduate classes!