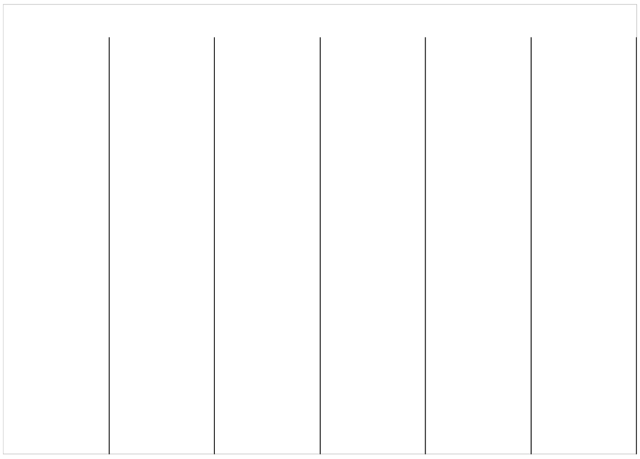
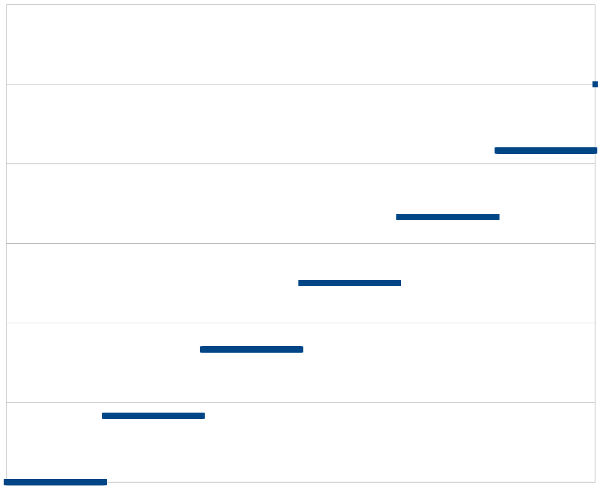
A vast international conspiracy!

Adam Atkinson (ghira@mistral.co.uk)

There is a conspiracy within mathematics. It spans many decades, and at least a few countries. How many decades and which countries? Perhaps you can help me find out! It's also possible that those of you who do outreach type activities will agree with me the topic suppressed by this conspiracy is odd enough to be of some value for outreach purposes.

Probability courses at high school and early undergraduate level usually mention discrete distributions (which have probability mass functions) and continuous distributions (which have probability density functions).

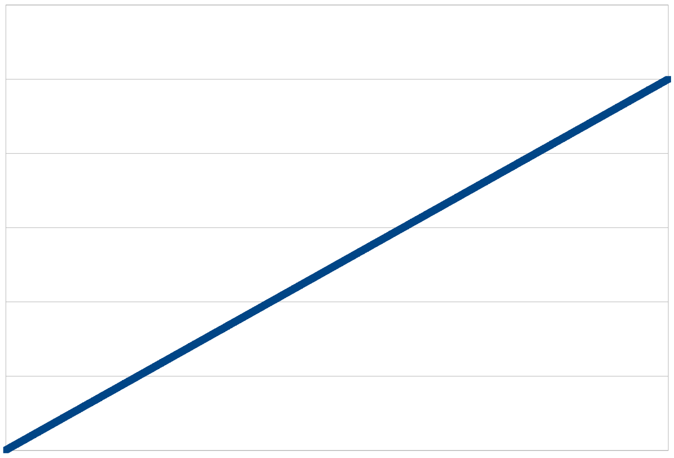
The mass function and cumulative distribution function for a fair die would look something like these

The CDF has jumpy bits. And flat bits.

It is worth admitting that discrete distributions can look a lot stranger than this. A distribution where all rationals between 0 and 1, or indeed all rationals, have positive probability is a discrete distribution. The mass function and CDF would both be quite hard to draw.

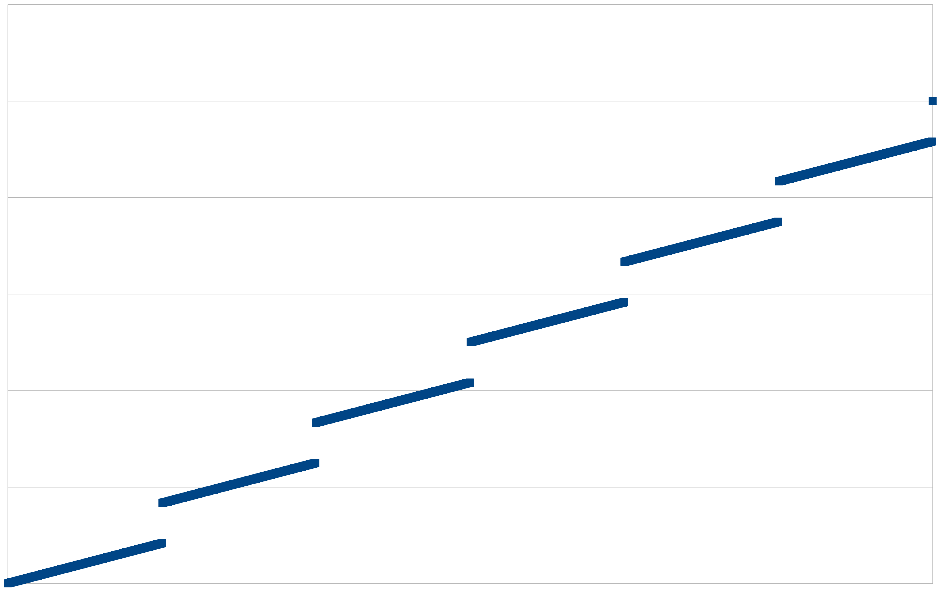
The density function and cumulative distribution function for a spinner producing real numbers from 0 to 6 would look something like this:

This particular CDF has a slopy bit with flat bits on either site. In principle it could have a mixture of slopy bits and flat bits. But no jumpy bits.

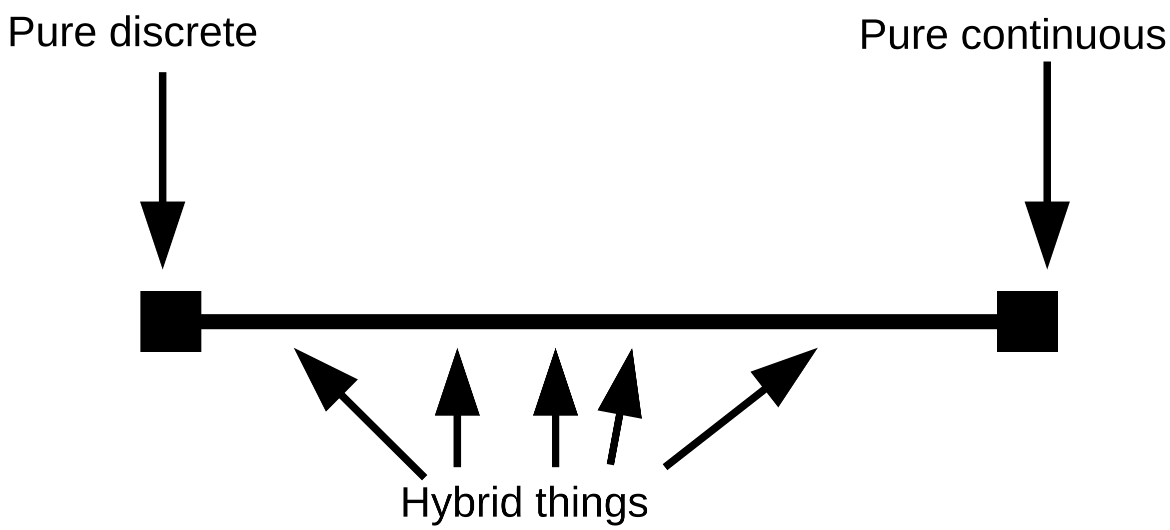
Distributions which are neither discrete nor continuous do often get mentioned in passing. An example would be a machine which generates integers from 1 to 6 50% of the time and real numbers from 0 to 6 50% of the time. The distribution of numbers emitted by this machine is clearly neither discrete nor continuous, and has neither a mass function nor a density function. It does still have a cumulative distribution function. However, this isn't really anything *very* new, just a mixture of two things we know about.

The cumulative distribution function here looks like this.

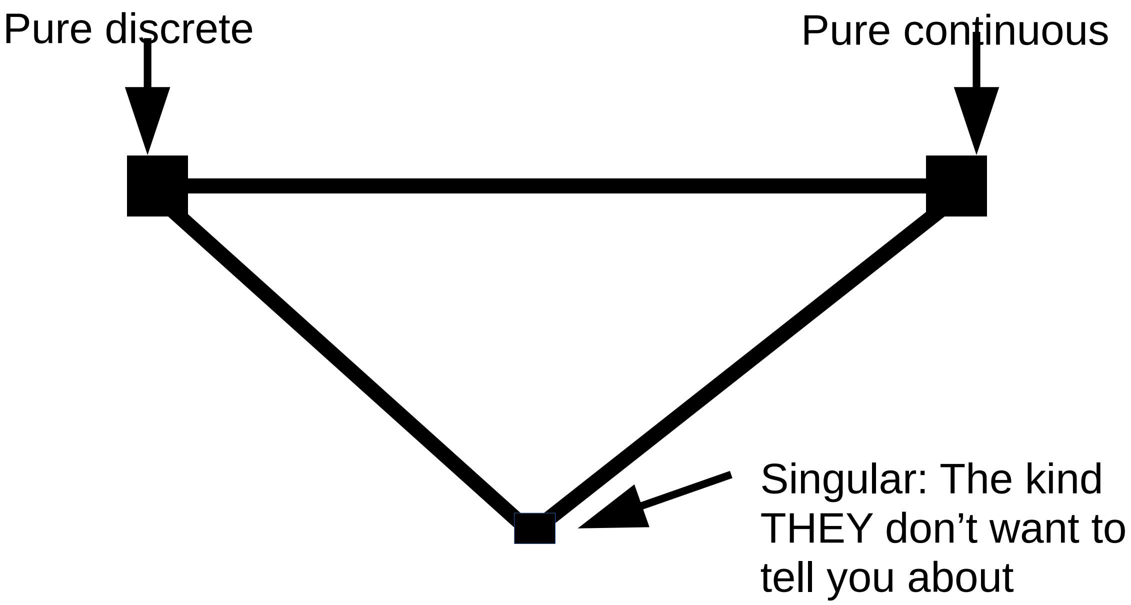


It has both jumpy bits and slopy bits. And flat bits to either site of the part of the graph we are looking at. In general, a hybrid made from a discrete and a continuous distribution could have a mix of jumpy, slopy and flat bits. And instead of a 50/50 split we could have weight the hybrid as we pleased towards the discrete or continuous components.

At this point we are perhaps left with the impression that probability distributions are on a spectrum a bit like this with pure discrete and pure continuous distributions at the ends.



The conspiracy involves a *third* basic kind of distribution. A general probability distribution is a linear combination of *three* components: discrete, continuous and singular.



Since we know about the first two already, at this point we want to know more about pure singular distributions.

Look for them in probability textbooks and see what you find. Dr Rachel Traylor (@mathpocalypse on Twitter, website http://www.themathcitadel.com/) and I looked in quite a few from the 30s to the present day and what we found could be divided into several classes:

- nothing
- There is another possibility but we won’t tell you what it is
- There is another possibility but it’s rather contrived so we won’t tell you what it is
- There is another possibility but it rarely turns up in applications so we won’t tell you what it is
- Some sort of description with or without a name but no examples.
- Description and an example. Always the *same* example.

Of course the final item on this list is not conspiratorial, but the others are much commoner and between them do leave you wondering half-seriously if there is in fact a conspiracy of some kind.

If you go looking for this, bear in mind that looking for “singular” or “distributions: singular” in the index probably isn’t enough. If they’re mentioned at all it’s often not by name, so you need to hunt around in the bits of books where discrete and continuous distributions are defined to look for passing references to other possibly nameless things. I am not providing bibliographic references partly because I’m not sure how you cite the absence of something.

It seems I was very lucky when younger to have been exposed to “Probability: An Introduction” by Grimmett and Welsh. It has several pages on singular distributions, and includes the usual example.

I asked quite a few friends (with degrees and/or PhDs in maths and related fields) and other contacts what they knew about singular distributions and if they knew of more than one actual example. Most people I asked had never heard of them. A few said “Of course I’ve heard of them. Hasn’t everyone?” But if I recall correctly no-one knew of a second concrete example of one.

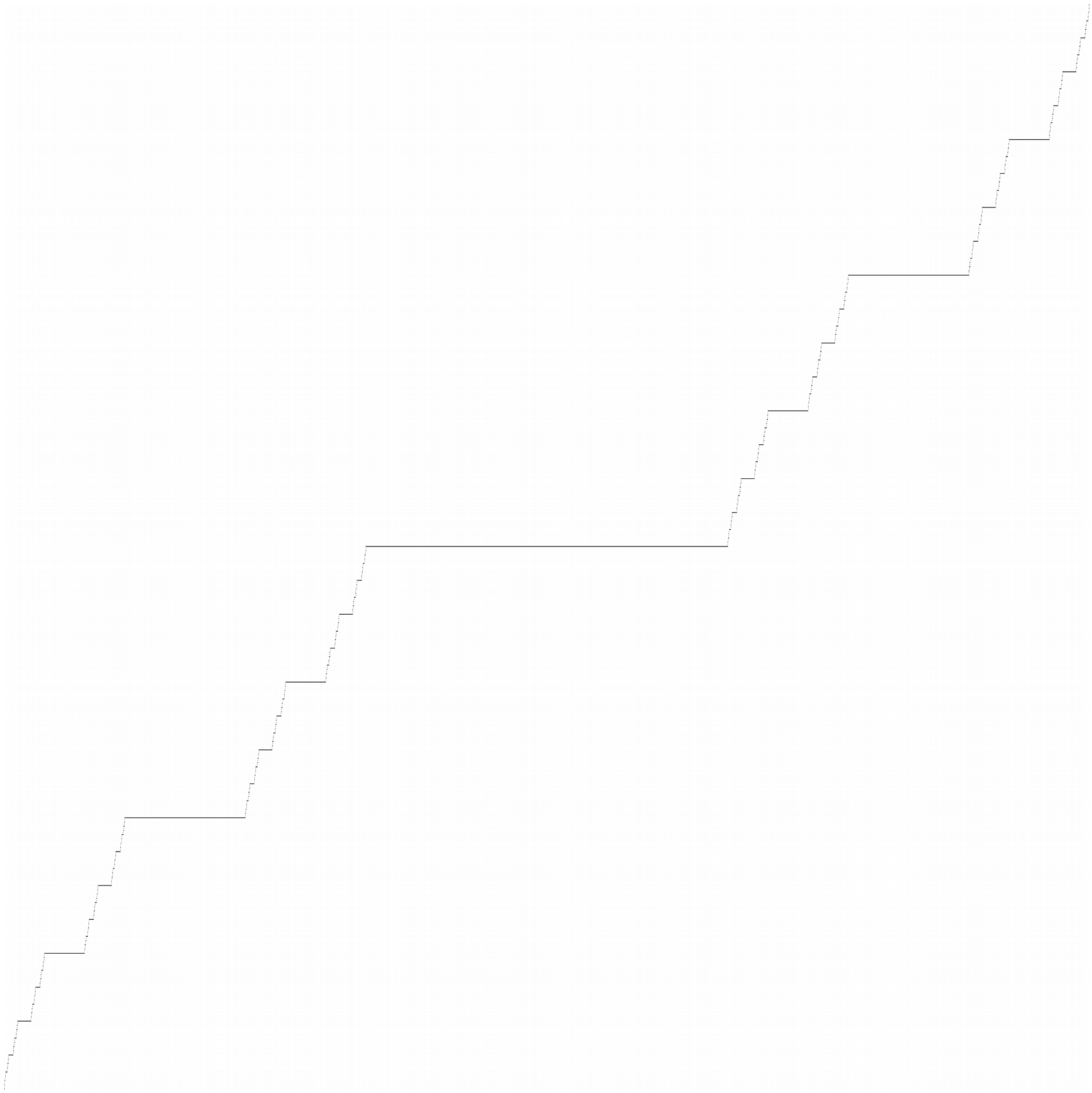
Rachel Traylor and I independently thought that of course “Counterexamples in Probability” by Stoyanov would have more examples. It’s a book of strange examples! And we were independently somewhat surprised and disappointed to learn that it actually says they exist but “They will not be treated here. The interested reader is referred to the books by...”. So singular distributions are too obscure even for a book of counterexamples?

Here is the example which always seems to be given if one is given at all:

Take a fair coin (whose tosses are independent). Throw it infinitely many times, one for each positive integer. As you go, accumulate a reward by taking $2/3^n$ pounds (or dollars, or Euros) if the nth result was heads. Your total reward can be between 0 (all tails) and 1 (all heads).

What is the distribution of the possible total rewards? If the CDF can’t have jumpy bits (since those belong to discrete distributions) or slopy bits (those belong to continuous distributions), what can it possibly look like? Of course we can have flat bits.

Here is the cumulative distribution function (between (0,0) and (1,1)):



It’s the devil’s staircase function! Make it flat with value ½ in the middle third of the unit interval. Then flat with values ¼ and ¾ in the middle thirds of the remaining intervals, and so on forever.

This distribution is called the Cantor distribution. The CDF is differentiable almost everywhere with derivative 0. It’s continuous. It has neither mass nor density. And it cannot be expressed as a linear combination of CDFs that do have them. Quite a few people I asked about singular distributions said that of course they were familiar with this function but that it had never been suggested to them that it could be used as the CDF of a probability distribution.

Since the CDF is continuous, there are no jumpy bits. There aren’t any slopy bits either since it’s flat almost everywhere. All the action takes place on an uncountable set of measure 0. The Cantor set, indeed.

Using the coin-tossing origin of this particular example we can see that the mean is ½ and that the variance is 1/8. (Exercise for the reader!)

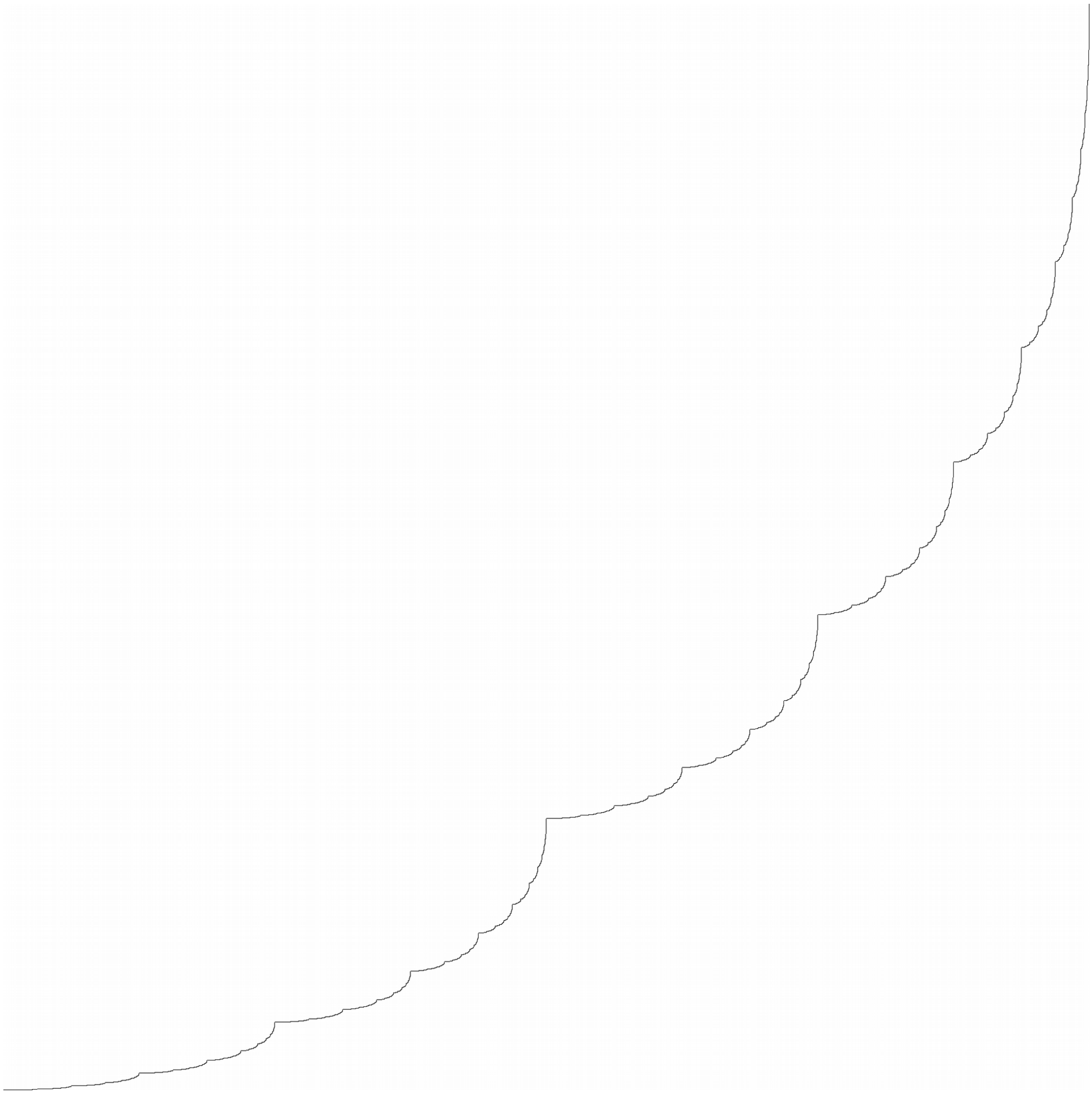
http://www.ghira.mistral.co.uk/cantor.mp4 shows a video of a family of related distributions where instead of middle thirds of intervals being flat we use middles covering a range of sizes. (The gaps that appear in the video are not really there, of course.)

We could also move the flat bits left or right and/or up and down. Or use several “middles” at each level. e.g. value 1/3 from 1/5 to 2/5 and 2/3 from 3/5 to 4/5.

Clearly we could also stretch/shrink/move the range of one of the distributions to cover a finite interval other than [0,1]. We could also stretch the x and/or y axes non-linearly, possibly even to cover the whole of the real line.

So, do all singular distributions look more or less like this? Lots of flat bits and some jiggery-pokery?

No, they do not. If you look around enough for a second example, you will find this one in various places. Here is the CDF:



Again it is differentiable almost everywhere with derivative 0. But this time it is also strictly increasing. So not only do we have no jumpy bits and no slopy bits, we no longer have any proper “flat bits” either. This seems much more disturbing than the Cantor distribution so you’d think if textbooks were going to give one example it should be this instead.

How do we calculate the mean, variance or other moments of a thing like this? This is a question which has turned up on Stack Exchange: https://math.stackexchange.com/questions/737148/how-we-compute-expectation-of-a-singular-random-variable

However, at least the Cantor distribution can be presented as something that arises from a coin-tossing exercise. In this case we’re just saying “Get a load of this!” and presenting a CDF. To be sure “Get a load of this!” is the way a lot of discrete and continuous examples are presented.

Actually the function we see here does arise in a classic probability problem, but not as a CDF. (Again, some people told me they knew of this function but had never seen it proposed as a CDF.)

This graph is one distribution from a family. Here is a video of several hundred members of this family with a slowly varying parameter (the value of f(1/2), as it happens). http://www.ghira.mistral.co.uk/boldplay.mp4

The way to obtain this function is as follows: define $f(0)=0$, $f(1)=1$, $f(1/2)=p$. Then for all other values in (0,1) require that $f(x)=pf(2x)$ if $x<=1/2$ or $p+(1-p)f(2x-1)$ if x is $>=1/2$.

So now we have two examples, or two families of examples. Is there a third example or family that turns up in books or articles? If we want to construct new singular distributions, how could we do it?

Three methods spring to mind:

1) Stretching or distorting examples we already have, possibly non-linearly, as previously mentioned.

2) Cutting and pasting pieces of examples we already have and using them to replace parts of the CDF of any distribution we like. If we replace the section of a CDF between $(x,f(x))$ and $(y,f(y))$ with a scaled/stretched piece of a singular distribution, and do the same with other sections of the CDF as we please until we have replaced everything, we can create a new singular distribution

3) Fractal interpolation. The graphs and videos seen so far were both produced with software I use to draw fractals. These particular fractals happen to be the graphs of continuous non-decreasing functions on the real line. The graph of the canonical Cantor distribution can be obtained by starting with the four points (0,0), (1/3, ½), (2/3, ½) and (1,1). The “boldplay” distributions I show are obtained by starting with (0,0), (½,p) and (1,1), with p strictly between 0 and 0.5.

More generally we could use fractal interpolation with (0,0), (1,1) and any finite number of intermediate points. Clearly since our function has to be nondecreasing as the x values of our points increase the y values can’t decrease. We should also avoid having points with the same x value and distinct y values. Given that the boldplay distribution with p=0 is discrete, having the same y value more than once can also cause problems though with the Cantor distribution it works out ok. Of course if all our points are on the diagonal line from (0,0) to (1,1) we will get a continuous distribution.

It has been suggested to me that singular distributions are not covered in relatively elementary courses because they are obscure, contrived and useless, and that more advanced work is so general that all distributions are covered so bothering to single out pure singular distributions would be peculiar. I don't know how reasonable or accurate this is.

I am interested in hearing more about how far in time and space the conspiracy spreads. What, if anything, did you "do" about singular distributions at school or university? If you teach probability, what if anything do you tell your students and why or why not? If you've seen a singular distribution in "real life" I'd love to hear about it. If you know of any examples which are interestingly different from those mentioned here, likewise. How can I create singular distributions on the whole of R other than by nonlinear stretching or infinite copy-and-pasting? Is there a fairly natural example of a singular distribution on the whole of R which can be defined directly?

The greatest amount of information I've found about singular distributions is in "Characteristic Functions" by Lukacs which is clearly not an elementary probability textbook.