The Measurement of Statistical Evidence
Lecture 2 - part 2

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/sta4522/STA4522.html

2021

What is randomness?

- recommended Sugita, H. (2018) Probability and Random Number, World Scientific, Chapter 2 a reasonably accessible treatment

- this issue was resolved in the 1960's by Chaitin, Kolmogorov and Solomonoff

- so what role does randomness have in probability and statistics?

- the concept of randomness is intimately connected with rigorizing what it means for a function f to be "computable" by a computer

- f is computable if there is a program to evaluate it and, since all inputs and outputs correspond to a finite binary sequences, we can restrict attention to the set $\mathcal{F}$ of all functions $f : \mathbb{N}_0 \to \mathbb{N}_0$

- $\mathcal{F}$ is uncountable (if countable, then can write $\{f_1, f_2, \ldots\}$ and define f by $f(i) = 0$ if $f_i(i) \neq 0$ and $f(i) = 1$ if $f_i(i) = 0$) so $f \in \mathcal{F}$ but $f \neq f_i$ for any i)

- since every program (algorithm) corresponds to a finite binary sequence, and the set of all finite binary sequences is countable, this implies that the set $\mathcal{F}$ of all computable functions (also called recursive functions) is countable

Kolmogorov Complexity (a sketch)

- put $\{0,1\}^*$ = set of all finite sequences of 0's and 1's obtained from elements of $\mathbb{N}_0$ by binary expansion with highest order bit equal to 1, let the empty sequence correspond to $0 \in \mathbb{N}_0$, and consider $\mathbb{N}_0$ and $\{0,1\}^*$ as identified

- so $0 \equiv ()$, $1 = 1 \cdot 2^0 \equiv (1), 2 = 0 \cdot 2^0 + 1 \cdot 2^1 \equiv (0,1), 3 = 1 \cdot 2^0 + 1 \cdot 2^1 \equiv (1,1), 4 = 0 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 \equiv (0,0,1)$, etc.

- **note** - a *recursive function* (see reference for precise definition) on $\mathbb{N}_0$ is a function that can be constructed from some basic functions and an operation called minimization and a *partial function* (as opposed to a *total function*) f on $\mathbb{N}_0$ means that it may only defined for some elements of $\mathbb{N}_0$

- let $l(x)$ = length of $x \in \{0,1\}^*$ (or $x \in \mathbb{N}_0$) and note for $x \in \mathbb{N}_0$, then $l(x) \leq \log_2 x + 1$ since $x = 2^{\log_2 x}$

Definition If $A : \{0,1\}^* \to \{0,1\}^*$ is a partial recursive function, when considered as a function $\mathbb{N}_0 \to \mathbb{N}_0$, then it is called an *algorithm*. The *computational complexity under algorithm* A of $x \in \{0,1\}^*$ is defined by

$$K_A(x) = \min\{l(q) : q \in \{0,1\}^*, A(q) = x\}$$

when $\{q \in \{0,1\}^*, A(q) = x\} \neq \phi$ and $K_A(x) = \infty$ otherwise.

Theorem There exists algorithm A_0 such that for any algorithm A there is a constant $c_{A_0,A}$ s.t. $K_{A_0}(x) \leq K_A(x) + c_{A_0,A}$.

- such an A_0 is called a *universal algorithm* (not unique but the Theorem also applies to two universal algorithms and their absolute difference is bounded by a constant)

- so when $K_{A_0}(x)$ is big, say much bigger than $c_{A_0,A}$ and bigger than $K_A(x)$, then $(K_{A_0}(x) - K_A(x))/K_{A_0}(x)$ is small

- for fixed A_0, $K(x) = K_{A_0}(x)$ is called the *Kolmogorov complexity* of x

- if a different universal algorithm is used, the absolute difference in the Kolmogorov complexities is bounded by a fixed constant

Theorem
(i) There exists constant $c > 0$ s.t. $K(x) \leq n + c$ for every $x \in (\{0,1\}^n)^*$ and $n \in \mathbb{N}_0$. So $K : \{0,1\}^* \rightarrow \mathbb{N}_0$ is a total function.
(ii) If $n > c' > 0$ then

$$\#\{x \in (\{0,1\}^n)^* : K(x) > n - c'\} > 2^n(1 - 2^{-c'}).$$

- $x \in (\{0,1\}^n)^*$ is called *random* if $K(x) \approx n$
- this "complexity" measure of x is a measure of the randomness of the sequence, e.g. $(0,1,0,1,0,1,\ldots,0,1)$ is not random
- for large n, Theorem (ii) implies most elements of $(\{0,1\}^n)^*$ are random

Example a computer program has computed 31.4 trillion decimal digits of π, or approximately $\log_2(10^{31.4\times 10^{12}}) = 1.04 \times 10^{14}$ bits, and the program for this is considerably shorter so this approximation to π is not random
- so far, although most elements of $\{0,1\}^*$ are random there is no known example of such a sequence

Theorem K is not a computable function (there is no program to compute it guaranteed to work).

- what this means is that there is no computable test for randomness

- implications for statistics: there is no test for randomness

- so what do current tests for randomness test? independent and identically distributed

Example *Champernowne's sequence*

- consider the following sequence $(x_1, ..., x_n) = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 0, 1, 1, 1, 2, 1, 3, 1, 4, ..., \cdot)$ and subject this sequence to a rest that the sequence is $i.i.d.$ from a uniform distribution on $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- if n is large enough the sequence will pass the test but it is not random

- recall, we stated earlier that the correct way to collect data was through *randomization* and, in particular, that this made the relative frequency distribution f_X suitable for assigning beliefs concerning the values taken by measurement X, namely, our belief that $X(\omega) \in A \subset \mathcal{X}$ is measured by

$$P(A) = \sum_{x \in A} f_X(x)$$

- why?

- my answer

> *Physical randomization corresponds to collecting the data in such a way that interested parties have absolutely no influence over the outcomes and, because of this, we can assert that the data is objective.*

- there are physical systems, like coin tossing, drawing chips from a bowl, that we **believe,** when performed appropriately, cannot be controlled or manipulated and so we accept these as random systems and use them to randomize

- so randomness has nothing to do with probability, which measures belief, but it plays a key role in ensuring that the data is objective

Cox's Theorem

- R. Cox (1946) attempted to characterize probability via a set of simple axioms in the sense that, if we accept the axioms, then the correct way to measure beliefs is via a probability measure P, at least up to a 1-1 transformation of P

- the attractive aspect of this approach is that it did not involve utilities or relative frequencies rather it was more based on the logical properties we would want such a measure to have

- such a theorem was proved by Cox but a flaw in the proof was discovered in 1999

- this was fixed in 2009 but the modification is not very appealing

- so a general open problem in this area is to find a development similar to Cox's that is also simple and appealing

- see the text for more details and references