Think Stats

A Conceptual Introduction to Statistics

for the Skeptical, the Pessimistic, and the Mildly Disturbed

G. A. Reynolds

February 28, 2014

Contents

1 Introduction

What is statistics?

> Modern statistics provides a quantitative technology for empirical science; it is a logic and methodology for the measurement of uncertainty and for an examination of the consequences of that uncertainty in the planning and interpretation of experimentation and observation.(Stigler, 1986, p. 1)

"If all sciences require measurement–and statistics is the logic of measurement– it follows that the history of statistics can encompass the history of all of science.(Stigler, 1986, p. 2)"

2

2 Sets 'n Stuff

We use the Z Specification notation (13568:2002, 2002).

Sets Membership, subset; family of sets

Relations

Functions

Sequences

Multisets

2.1 Sets

Notation: extension v. comprehension

2.2 Relations

2.3 Functions

2.3.1 Definition

Function Informally, a function is a set of ordered pairs. The Z specification says "A function is a particular form of relation, where each domain element has only one corresponding range element."(13568:2002, 2002, p. 9)

Function Extension Since a function is a kind of set, it can be defined by explicitly listing its extension–all of its elements. For example, the function f that maps each integer between 1 and 3 to itself can be expressed by writing out the complete list of its elements: $f = \{(1,1),(2,2),(3,3)\}$.

Function Comprehension A second way of defining a function is to express the "rule" that determines the elements it contains, without listing them explicitly. For example "the function that maps every number to itself" defines the identity function. Z provides two ways of doing this, one using standard set comprehension notation, the other using "function construction" notation. See below.

2.3.2 Notation

The Z notation supports several ways of representing a function. A function extension expression may use ordered pair notation or maplet notation. The following example illustrates two ways to define the function that maps each integer between 1 and 3, inclusive, to its double.

$$f = \{(1, 2), (2, 4), (3, 6)\} \tag{2.3.1}$$
$$= \{1 \mapsto 2, 2 \mapsto 4, 3 \mapsto 6\} \tag{2.3.2}$$

More generally:

$$\langle e_1, \ldots e_n \rangle = \{(1, e_1), \ldots (3, e_n)\} \tag{2.3.3}$$
$$= \{1 \mapsto e_1, \ldots 3 \mapsto e_n\} \tag{2.3.4}$$

Z also supports two ways to define a function intensionally – in terms of a property rather than an explicit list of elements.

Remark 1 *FIXME: set comprehension expression for functions:*

$$f = \{x, y \mid y = 2x \bullet (x, y)\} \tag{2.3.5}$$
$$= \{x, y \mid y = 2x \bullet x \mapsto y\} \tag{2.3.6}$$

Function construction notation...

$$f = \{\lambda \tag{2.3.7}$$

2.4 Sequences

2.4.1 Definition

Sequence Informally, a sequence is an ordered set. The Z specification says "A sequence is a particular form of function, where the domain elements are all the natural numbers from 1 to the length of the sequence."

2.4.2 Notation

The Z notation uses angle brackets to form a sequence expression:

$$\langle e_1, \ldots e_n \rangle = \{(1, e_1), \ldots (3, e_n)\} \tag{2.4.1}$$
$$= \{1 \mapsto e_1, \ldots 3 \mapsto e_n\} \tag{2.4.2}$$

4

2.5 Multisets

We follow (Singh et al., 2007), with some modifications.

2.5.1 Definitions

Multiset

Element

Carrier The carrier of an mset is the set from which its elements is drawn.

Generator The generators of an mset are the elements of its carrier set.

Multiplcity The multiplicity of an element of an mset is the number of times it "occurs" ("appears", etc.).

Cardinality The cardinality (size) of an mset is the sum of the multiplicities of its elements. The cardinality of the carrier of an mset is the number of elements it contains.

2.5.2 Notation

The mset containing one a, two b, and three c

$$M = \{(a, 1), (b, 2), (c, 3)\} \tag{2.5.1}$$

can be written as follows:

Multiplicative notation The following are equivalent:

- $[[a, b, b, c, c, c]]$
- $[a, b, b, c, c, c]$
- $[a, b, c]_{1,2,3}$
- $[a^1, b^2, c^3]$
- $[a1, b2, c3]$

Linear notation The following are equivalent:

- $[a] + 2[b] + 3[c]$

- $\{[a], 2[b], 3[c]\}$ Note that this style combines multiplicative nota-
  tion for mset elements with standard set notation ($\{\ldots\}$) for the
  mset itself.

One advantage of the linear notation is that it allows us to have non-
integral and non-positive multiplicities; for example, in $\{[a], -0.5[b], \pi[c]\}$
element a occurs once, b occurs -0.5 times, and c occurs π times.

2.6   Multisequences

A multiset is a set of ordered pairs, and is therefore unordered. Just as we
can extend the concept of set to form a sequence, we can extend the notion
of multiset to form a multisequence.

Multisequence A multisequence is a sequence of multiset elements.

We use the same angle bracket notation we use for set sequences:

$$\langle [a], 2[b], 3[c] \rangle = \{(1, (a, 1)), (2, (b, 2)), (3, (3, c))\} \tag{2.6.1}$$
$$= \{1 \mapsto (a, 1), 2 \mapsto (b, 2), 3 \mapsto (3, c)\} \tag{2.6.2}$$

Since multisets are unordered, we have

$$\{[a], 2[b], 3[c]\} = \{2[b], [a], 3[c]\} = \{3[c], 2[b], [a]\} = \ldots \tag{2.6.3}$$

Multisequences are ordered, so for example

$$\langle [a], 2[b], 3[c] \rangle \neq \langle 2[b], [a], 3[c] \rangle \tag{2.6.4}$$

since

$$\{(1, (a, 2), (2, (b, 2), (3, (c, 3)\} \neq \{(1, (b, 2)), (2, (a, 2)), (3, (c, 3))\} \tag{2.6.5}$$

alternatively

$$\{(1 \mapsto (a, 2), 2 \mapsto (b, 2), 3 \mapsto (c, 3)\} \neq \{(1 \mapsto (b, 2)), 2 \mapsto (a, 2)), 3 \mapsto (c, 3))\} \tag{2.6.6}$$

3 Series

4 Counting and Combining

(Berge, 1971)

What is combinatorics?

"Combinatorics can rightly be called the mathematics of counting. More specically, it is the mathematics of the enumeration, existence, construction, and optimization questions concerning nite sets. (Mazur, guided tour)"

> The concept of configuration can be made mathematically precise by defining it as a mapping of a set of objects into a finite abstract set with a given structure; for example, a permutation of n objects is a bijection of the set of n objects into the ordered set 1, 2, ..., n. Nevertheless, one is only interested in mappings satisfying certain constraints.

> Just as arithmetic deals with integers (with the standard operations), algebra deals with operations in general, analysis deals with functions, geometry deals with rigid shapes, and topology deals with continuity, so does combinatorics deal with configurations. Combinatorics counts, enu- merates,* examines, and investigates the existence of configurations with certain specified properties.

4.1 Counting Principles

Remark 2 *Most of these are just restatements of basic arithmetic, expressed in terms of doing things. Why bother? I suspect that thinking in terms of sequences of actions makes it easier to do combinatorics. Also, these ideas only seem to be articulated as principles in elementary texts for e.g. high school algebra.*

Remark 3 *But isn't combinatorics just the science of counting?*

Principle 1 (Fundamental Principle of Counting)

Principle of addition : if there are a ways of doing one thing and b ways of doing another, and we cannot do both, then there are $a + b$ ways

to choose one thing to do. This is just a restatement of a set-theoretic definition of addition in terms of union of sets.

In combinatoric texts something like this is more typical:

Addition Rule. If A and B are finite, disjoint sets, then $A \cup B$ is finite and $\mid A \cup B \mid = \mid A \mid + \mid B \mid$.

Principle of multiplication : if there are a ways of doing one thing and b ways of doing another, then there are $a \cdot b$ ways of doing both. This is a restatement of a set-theoretic definition of addition in terms of cartesian products.

5 Computability and Decideability

Remark 4 *Why important for intro to stats? Mainly for historical reasons; both effective procedure and axiom of choice emerged at roughly the same time. Also, an understanding of effective proc sharpens understanding of choice and randomness.*

5.1 Effective Procedure

6 Choice and Chance

6.1 The Axiom of Choice

The Axiom of Choice is of enormous important in mathematics generally. Statistics is no exception; the significance of the axiom will become especially apparent when we discuss the concept of random sample.

The principle of set theory known as the Axiom of Choice has been hailed as probably the most interesting and, in spite of its late appearance, the most discussed axiom of mathematics, second only to Euclid's axiom of parallels which was introduced more than two thousand years ago (Fraenkel, Bar-Hillel & Levy 1973, II.4). The fulsomeness (*sic*) of this description might lead those unfamiliar with the axiom to expect it to be as startling as, say, the Principle of the Constancy of the Velocity of Light or the Heisenberg Uncertainty Principle. But in fact the Axiom

of Choice as it is usually stated appears humdrum, even self-evident. For it amounts to nothing more than the claim that, given any collection of mutually disjoint nonempty sets, it is possible to assemble a new set a transversal or choice set containing exactly one element from each member of the given collection. Nevertheless, this seemingly innocuous principle has far-reaching mathematical consequences many indispensable, some startling and has come to figure prominently in discussions on the foundations of mathematics. It (or its equivalents) have been employed in countless mathematical papers, and a number of monographs have been exclusively devoted to it.(Bell, 2013)

Often stated in terms of choice functions.
Variants:
AC1: Any collection of nonempty sets has a choice function."
AC2: Any indexed collection of sets has a choice function.
Or relations:

Axiom 1 (Axiom of Choice) *For every family $\mathcal{F}$ of nonempty disjoint sets there exists a selector, that is, a set S that intersects every $F \in \mathcal{F}$ in precisely one point.(Ciesielski, 1997)*

Transversal: In a 1908 paper Zermelo introduced a modified form of AC. Let us call a transversal (or choice set) for a family of sets H any subset T H for which each intersection T X for X H has exactly one element. As a very simple example, let H = 0, 1, 2, 3. Then H has the two transversals 0, 1, 2 and 0, 1, 3. A more substantial example is afforded by letting H be the collection of all lines in the Euclidean plane parallel to the x-axis. Then the set T of points on the y-axis is a transversal for H.

So we have choice functions and choice sets.

"Let us call Zermelo's 1908 formulation the combinatorial axiom of choice:

CAC: Any collection of mutually disjoint nonempty sets has a transversal." (bell)

The problem:

It is to be noted that AC1 and CAC for finite collections of sets are both provable (by induction) in the usual set theories. But in the case of an infinite collection, even when each of its members

is finite, the question of the existence of a choice function or a transversal is problematic[4]. For example, as already mentioned, it is easy to come up with a choice function for the collection of pairs of real numbers (simply choose the smaller element of each pair). But it is by no means obvious how to produce a choice function for the collection of pairs of arbitrary sets of real numbers.(Bell, 2013)

Footnote:

The difficulty here is amusingly illustrated by an anecdote due to Bertrand Russell. A millionaire possesses an infinite number of pairs of shoes, and an infinite number of pairs of socks. One day, in a fit of eccentricity, the millionaire summons his valet and asks him to select one shoe from each pair. When the valet, accustomed to receiving precise instructions, asks for details as to how to perform the selection, the millionaire suggests that the left shoe be chosen from each pair. Next day the millionaire proposes to the valet that he select one sock from each pair. When asked as to how this operation is to be carried out, the millionaire is at a loss for a reply, since, unlike shoes, there is no intrinsic way of distinguishing one sock of a pair from the other. In other words, the selection of the socks must be truly arbitrary.

The axiom of choice and probability (randomness) are different concepts. See http://math.stackexchange.com/questions/29381/picking-from-an-uncountable-set-axiom-of-choice

6.2   Chance and Randomness

See (Eagle, 2014)

Indeterminacy, disorder, chaos, stochastic process, etc.

chance of a process, randomness of its product

Chance: physical; randomness: mathematical?

"It is safest, therefore, to conclude that chance and randomness, while they overlap in many cases, are separate concepts."(Eagle, 2014)

Process v. Product concepts

"Of course the terminology in common usage is somewhat slippery; it's not clear, for example, whether to count random sampling as a product

notion, because of the connection with randomness, or as a process notion, because sampling is a process."

> The upshot of this discussion is that chance is a process notion, rather than being entirely determined by features of the outcome to which the surface grammar of chance ascriptions assigns the chance. For if there can be a single-case chance of for a coin to land heads on a toss even if there is only one actual toss, and it lands tails, then surely the chance cannot be fixed by properties of the outcome lands heads, as that outcome does not exist.[2] The chance must rather grounded in features of the process that can produce the outcome: the coin-tossing trial, including the mass distribution of the coin and the details of how it is tossed, in this case, plus the background conditions and laws that govern the trial. Whether or not an event happens by chance is a feature of the process that produced it, not the event itself.(Eagle, 2014)

> a process conception of randomness makes nonsense of some obvious uses of random to characterise an entire collection of outcomes of a given repeated process. This is the sense in which a random sample is random: it is an unbiased representation of the population from which it is drawnand that is a property of the entire sample, not each individual member. While many random samples will be drawn using a random process, they need not be....To be sure that our sample is random, we may wish to use random numbers to decide whether to include a given individual in the sample; to that end, large tables of random digits have been produced, displaying no order or pattern (RAND Corporation 1955). This other conception of randomness, as attaching primarily to collections of outcomes, has been termed product randomness.(eagle)

> If the actual process that generate the sequences are perfectly deterministic, it may be that a typical product of that process is not random. But we are rather concerned to characterise which of all the possible sequences produced by any process whatsoever are random, and it seems clear that most of the ways an infinite sequence might be produced, and hence most of the sequences so produced, will be random.(eagle)

11

...concentrate on the sequence of outcomes as independently given mathematical entities, rather than as the products of a large number of independent Bernoulli trials...

Goal: devise mathematics to "capture the intuitive notion of randomness."

Compare logicians' attempts to capture the intuitive notion of logical consequence.

Howson and Urbach (1993: 324) that it seems highly doubtful that there is anything like a unique notion of randomness there to be explicated.

Intuitions about randomness:

- a property of a sequence

- indeterminism

- epistemic randomness

See also https://www.cs.auckland.ac.nz/ chaitin/sciamer.html

7 Infinity and the Limit Concept

8 Measure and Integration

Remark 5 *Why this? Mainly just as a notational convenience. Even if we target an audience with minimal math, it is useful to have $\int$ (or $\overset{+}{\lambda}$) to indicate the area under a curve. And that is critical to the concept of probability of a continuous random variable, which is one of the main concepts we want to get across.*

Furthermore the basic concepts are not that difficult: differentiation as the limit of a ratio of differences, integration is the limit of a sum of products. The details of how one might actually do this may be daunting for many readers, but the basic concepts are quite simple and intuitive. Especially with lambda notation.

Remark 6 *What about differentiation?*

9 Measurement and Error

Remark 7 *The previous section discuss the mathematical concept of measure. This section discusses theories of empirical measurement.*

10 Log and Exp

Remark 8 *The purpose here is not to define log and exp per se, but to show how they are graphed, so that later the graphs of probability distributions like the normal curve are intelligible.*

11 Random Variables

Random selection devices

12 Probability Distributions

13 Sampling

Appendices

A Bibliography

[1] ISO/IEC 13568:2002. *Information technology – Z formal specification notation – Syntax, type system and semantics.* Geneva: ISO, 2002. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=21573 (visited on 02/28/2014) (cit. on p. 3).

[2] John L. Bell. "The Axiom of Choice". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Winter 2013. 2013. URL: http://plato.stanford.edu/archives/win2013/entries/axiom-choice/ (visited on 02/28/2014) (cit. on pp. 9, 10).

[3] C Berge. *Principles of Combinatorics.* Academic Press, Apr. 1971 (cit. on p. 7).

[4] Krzysztof Ciesielski. *Set Theory for the Working Mathematician.* London Mathematical Society student texts 39. Cambridge ; New York: Cambridge University Press, 1997 (cit. on p. 9).

[5] Antony Eagle. "Chance versus Randomness". In: *The Stanford Encyclopedia of Philosophy.* Ed. by Edward N. Zalta. Spring 2014. 2014. URL: http://plato.stanford.edu/archives/spr2014/entries/chance-randomness/ (visited on 02/28/2014) (cit. on pp. 10, 11).

[6] D. Singh et al. "An Overview of the Applications of Multisets". In: *Novi Sad J. Math.* 37.2 (2007), pp. 73–92 (cit. on p. 5).

[7] Stephen M Stigler. *The History of Statistics: the Measurement of Uncertainty Before 1900.* English. Cambridge, Mass.: Belknap Press of Harvard University Press, 1986 (cit. on p. 2).