Mathematics for Policy and Planning Science

Stephen Turnbull

Graduate School of Systems and Information

Lecture 1: April 18, 2018

Abstract

Introduction to Mathematics for Policy and Planning Science (1st half): the course, the instructor, and the field. Introduction to counting and probability.

The Times They Are A-Changin'

Come gather 'round people, where ever you roam
And admit that the waters around you have grown
And accept it that soon, you'll be drenched to the bone
If your time to you is worth saving
Then you'd better start swimming or you'll sink like a stone
For the times, they are a-changin'.

- Lost decades, Chinese economic growth, Lehmann shock, Brexit, Trump: the times are changing.

We Need to Change with the Times

- It seems like everybody in Japan hopes for a return to the age of *monozukuri* and lifetime employment. Many professors in *Shakō* teach about *machizukuri*, which is pretty close. Unfortunately, these days Japan has a lot more jobs pouring drinks than pouring concrete, even with earthquake disasters and the 2020 Tokyo Olympics.
- Those are still value-added areas in goods production where Japan can lead —but they probably won't lead to great increases in employment.
- To employ all her people at high wages, Japan needs to become a service economy. And not just services, but one focused on innovation in customized products—just as the U.S. has already done.
- China is not far behind Japan, Europe, and Korea on the same trajectory.

The road to tomorrow

- Many of you will prefer "traditional" paths to employment, with big manufacturing companies or the public sector. But I hope that many will turn to ventures and the service sector. Modern economies need leaders in innovation. And many say government and traditional mass production industries should innovate, too.

- To lead in innovation at the "global" level, one must engage in *service engineering*.
  - Use the web to enable (or for information services, to implement) customized products.
  - Theory-based implementation and marketing (not experimentation and *kaizen*) to take advantage of "market windows". *E.g.*, fast fashion (success), 3-CD boomboxes (failure), soft drinks (mixed).
  - Precise calculations to achieve larger margins in mature industries.

- "Engineering" is theory-based and quantitive, yet requires human expertise and judgment, and experimentation.

What is mathematics? Philosophy

- Prof. Saunders Mac Lane wrote a 456-page book to answer that question, in which he concluded:

  > Mathematics aims to understand, to manipulate, to develop, and to apply those aspects of the universe which are formal.

- By “formal aspects of the universe” Mac Lane means something like “those things which can be calculated.” But *calculation* itself is only a small part of Mathematics to Mac Lane.

- *Understanding* when a calculation is applicable, *manipulating* algorithms to produce new ways of calculating, and *developing* more applications where calculation is useful is the major part.

What is mathematics? An applied view

This quotation is from J. M. Chambers's *Software for Data Analysis: Programming with R*, and is about statistical software. But the same ideas apply to using mathematics as to programming.

> [Many readers] will have some experience ... with software for statistics, but will view their involvement as doing only what's absolutely necessary to "get the answers". This book will encourage moving on to think of the interaction with the software as an important and valuable part of your activity. You may feel inhibited by not having done much programming before. Don't be. Programming ... can be approached gradually, moving from easy and informal to more ambition projects. As you use [software], one of its strengths is its flexibility. (pp. vii–viii)

Substitute "mathematics" for "programming" and "software".

What is mathematics? In plain English

- When studying mathematics, you learn many calculations. But this can't be very important; computers calculate better and faster than you do, and you will use them in advanced study and at work.
- By studying mathematics, you learn to create *formal representations of models* which can be calculated. This is probably the most important thing to take away from this class: the idea that mathematics *is not real*, but rather *expresses a model of reality in a form convenient for calculation.*
- A consequence of "mathematics as model" is that you must always carefully consider whether you (or the computer!) are doing the right calculations.
- In other words, your contribution to a project is not the *numbers* you present, but rather the *choice* of calculations to do, and the assurance that those calculations are *appropriate.*
  - If you're the programmer or data engineer, you're also responsible for ensuring the calculations are *correct.*

Applications of mathematics: statistics

- Unlike most engineering disciplines, in Shako we don't generally use calculations to build things, although sometimes we do.
- Almost all of us need to use statistics. Statistics allow us to *estimate and test models with data*, by computing statistics from the data and comparing them with threshold values.
- A valid model may confirm a theory, or be used in prediction, or to discover variables which cannot be measured directly.

Using statistics: confirming gender differences

- Suppose you have data on a large group including men and women, and you'd like to know whether men are *in general* taller than women.
- Use the t (estimated difference of means divided by estimated standard error) statistic, of course.
- The meaning of "difference of means" is obvious. But what is the right denominator?
  - Compare the average of women to the men's average, using the standard error of the female sample (men's average as "standard").
  - Compare the average of men to the women's average, using the standard error of the male sample (women's average as "standard").
  - Take the difference, using the standard error of the whole sample.
- They're all wrong!

What are the assumed models?

- Using the standard error of *one* subsample requires assuming that the *estimated* mean of the other subsample is the *true* mean of the subsample.
- Using the standard error of the *whole* sample requires assuming that both populations have the same standard deviation.
- The model we want to use has men and women having *different* distributions in every way. That's probably not practical, and although no human being can have a negative height, the probability of a negative draw for a normal random variable with "reasonable" mean and standard deviation (say, 165cm *vs.* 15cm) is such that we'd need to try 10^{20} times to have a 50% chance of one negative draw.

  So, in practice we will use two normal distributions, but both mean and variance must be allowed to differ, and thus be estimated from sample data.

Brief course description

For the first half of this course, I will discuss statistics and its underlying mathematical ideas, as related to the goal of supporting research in Policy and Planning Sciences.

Goal Understanding of the basic ideas of data analysis using statistics, including the underlying quantitative tools (probability and linear algebra). Statistical models, descriptive statistics including factor analysis, hypothesis testing, and regression analysis will be introduced.

Overview of the Lectures We consider basic ideas about gathering, organizing, and analyzing data. Then we introduce simple statistical models and regression analysis. If time permits, I would like to mention recent developments like data mining techniques.

Prerequisites and Language

Prerequisites Although not absolutely necessary, for best results students should have taken college level calculus and linear algebra courses.

Language of Instruction I plan to lecture in English, and original course materials will generally be in English. I will accept and answer questions in Japanese to the extent possible (but my technical vocabulary is relatively weak; it's probably best to use English technical terms where possible).

- I hope to provide Japanese translations of some course materials. However, the Japanese may not be 100% reliable. Always check against the English.

Manual Calculation

- Calculation by hand will be a prominent feature of this class. N.B. “By hand” includes use of spreadsheets, but unfortunately I can’t permit that on examinations.
- Intended to improve your understanding and intuition about computations.
- Computers can do calculations more quickly, more accurately, and at far larger scale than any human is capable of, but they are a black box to any but expert software engineers. The “garbage in, garbage out” problem is as dangerous in statistical analysis as in any field of computation.

Computational Exercises

- Computational exercises will also be assigned.
- Intended to familiarize you with practical issues of input and output (*e.g.*, organization of data sets), and interpretation of common statistics.
- The intent is not to make you an expert at using computers. There are at least 4 major software packages in use in our faculty, and each has both advantages and disadvantages. You should choose one which is easy for you to use and adapted to your purpose, when that becomes clear (usually when a task such as a thesis is assigned to you). This is not a good time to specialize.

Resources

- Just about anything you need to know about the class will be on the class home page, `http://turnbull.sk.tsukuba.ac.jp/Teach/ShakoMath/`. If it's posted on my home page, "I didn't know (about the assignment, test, *etc.*)" will *not* be an acceptable excuse.
- The other important URL is my personal calendar, `http://turnbull.sk.tsukuba.ac.jp/schedule/`.
- There is no required *textbook*.

Recommended textbooks

- *Statistics*, by David Freedman, *et al.*. This book is an investment (*i.e.*, expensive but worth it). No advanced techniques, but many examples showing how to interpret statistics.
- *Principles and Practice of Structural Equation Modeling*, by Rex Kline. An advanced topic, but basic concepts of modeling are very well presented. Highly recommended for marketing and OB.
- *Introduction to Econometrics* by G. S. Maddala and Kajal Lahiri. A classic textbook, now in its fourth edition.
- *Introduction to Statistics and Econometrics*, by Takeshi Amemiya. The textbook presents some advanced concepts.

See the class home page for more information.

Why do we need proofs in Shako?

For want of a nail the shoe was lost,
for want of a shoe the horse was lost,
for want of a horse the knight was lost,
for want of a knight the battle was lost,
for want of a battle the kingdom was lost.
So a kingdom was lost—all for want of a nail.

- In a chain of reasoning, any missing step breaks the chain, causes the proof to fail, and can lead to erroneous conclusions.

A proof that $2 = 1$

Step 1: Let $a = b$.
Step 2: Then $a^2 = ab$,
Step 3: $a^2 + a^2 = a^2 + ab$,
Step 4: $2a^2 = a^2 + ab$,
Step 5: $2a^2 - 2ab = a^2 + ab - 2ab$,
Step 6: and $2a^2 - 2ab = a^2 - ab$.
Step 7: This can be written as $2(a^2 - ab) = 1(a^2 - ab)$,
Step 8: and cancelling the $(a^2 - ab)$ from both sides gives $2 = 1$.

`https://www.math.toronto.edu/mathnet/falseProofs/first1eq2.html`

Financial instruments

- For example, *martingale theory* can determine if the investment strategy "buy and hold until profit is X or greater, then sell" can be successful, *i.e.*, on average make more than zero profit.
- The answer for a simple random walk (which many experts consider to be a good model of securities *vs.* the market average) is "yes and no; really, no."
- Yes, there exists a "random time" T in the future when the profit is X or more, and therefore the *expected profit* is positive. But ...
- The expected value of T is not finite!
- The profit "often" reaches arbitrarily large *losses* before achieving the target profit of X!
- So, you see, really "no."

Path dependence and flexibility

This example comes from a real working paper by a famous professor.

- Path dependence means the outcome of a process depends on history. If history includes random events, one theory can predict multiple outcomes.
  - Compare the *central limit theorem*, which states that many long-run averages are always normally distributed with specific parameters.
- In game theory we typically formulate strategies *recursively*. That is
  - Each player $i = 1, 2, \ldots, n$ picks her first action unconditionally: $a_i^0 = \bar{a}_i^0$.
  - The second action is based on the first: $a_i^1 = f_i^1(a^0)$, where $a^t = (a_1^t, a_2^t, \ldots, a_n^t)$ for each period t.
  - The t-th period action is based on the entire history of preceding actions: $a_i^t = f_t^1(a^0, a^1, \ldots, a^{t-1})$.
- The set of possible *action combination histories* is huge, even in a repeated game with the same set of actions at each period (*e.g.*, the prisoner's dilemma, with "silence" or "confess"): $(A_1 \times A_2 \times \ldots \times A_n)^{\mathbb{N}}$.

Possible outcomes

- Consider a model where only the starting actions have a random component.
- Consider the trivial case of a *one* player with *two* actions at each stage.
- Then when the player starts with $a^0 = a$, we can predict the rest of her actions: $a^1 = f^1(a)$, $a^2 = f^2(a, f^1(a))$, $a^3 = f^3(a, f^1(a), f^2(a, f^1(a)))$, ...
- Since each a^t depends only on the initial choice of a, and there are only two possibilities, this extremely complex strategy where each action can flexibly depend on all of history can produce only two histories, one for each action in A^0!
- In a two-player game with two actions for each at the first, there are only four possible histories!
- Remember that in Nash equilibrium, each player "knows" (infers) the others' strategies. *This model cannot be used for a repeated game where hiding intentions over time is important.*

Counting

- The most basic math is counting.
  - We care about counting because we want to compare sizes of sets.
- The basic principle is that two sets with the same number of elements can be paired 1-1.
- So "2" is the invariant of couples: ie., sets where one is different from the other, and anything different from the one is the other.
  - "3" is more difficult to describe this way but it can be done.
- The natural numbers $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ constitute a *canonical set* for comparing sizes, and *labels* for those sizes.

Counting isn't trivial

- Counting seems easy, but you had to learn it (and it wasn't just a matter of learning names for numbers, the concept itself was learned).
- And there are important cases where it's not obvious what counting means.
- Mathematicians have discovered that there are (at least) two possible interpretations of number. One is ordinal, the other is cardinal. For finite numbers, they're the same, but for infinite numbers they're different!
- Here are some questions:
  - How many even numbers are there?
  - Is that number the same as for odd numbers?
  - How about if we leave out zero?

  Actually, they're all the same!
- Even more, the set of pairs of natural numbers, the set of rationals, and the set of natural numbers are all the same size.

Infinite subsets of $\mathbb{N}$

- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
- Pairing even numbers with odd numbers: $\mathrm{odd}(x) = x + 1$, for $x \in \{0, 2, 4, \ldots\}$.
- Pairing even numbers without 0 with odd numbers: $\mathrm{odd}(x) = x - 1$, for $x \in \{2, 4, 6, \ldots\}$.

Multidimensional infinities

Proof of equivalence of natural numbers, pairs of natural numbers, and rationals.

Different sizes of infinity

- Are there different sizes of infinity?
- The answer is ”yes”, and the idea of the proof isn't hard.
  - Assume a list of all real numbers between 0 and 1, represented as infinite decimals, say $\{a_0, a_1, a_2, \ldots\}$ (not necessarily in any order, except that of the list).
  - Such a list (not set!) has the same size as $\mathbb{N}$, which is infinite.
  - Now construct a decimal whose first digit is different from the first digit of a_0, whose second digit is different from the second digit of a_1, whose third digit is different from the third digit of a_2, and so on.
  - It's different from any number in the list, so the list is incomplete.
- Thus the set of real numbers between 0 and 1 is bigger (as an infinity) than the set of natural numbers.
- This matters to the theory of integral calculus.

Counting problems: Tumbling dice

- Sum of two dice has 11 values. Each outcome equally probable?
- No, "7" is most probable, "2" and "12" quite rare.
- Can we come up with a model with "something" equally probable, but the sum has the right distribution?
- Yes, by taking the product of 6-element sets.
- Homework: predict the distribution of *differences* of two dice.
- Homework: predict the distribution of *products* of two dice.
- Homework: predict the distribution of *quotients* of two dice.

Introduction to Probability

- *Probability* is a numerical assessment of "likelihood" of *events*.
- It obeys some convenient mathematical laws.
- It is useful in understanding "rational behavior":
  - If two actions A and B both result in the event "1000 yen profit," but with action A the result seems more likely (has *higher probability*) than with action B, it is "rational" to choose action A if only one of the two actions is possible.
  - In fact, this is an *axiom* of Laurence Savage's *theory of expected utility*.
  - In this sense, the economic theory of rationality is somewhat circular: we *define* rationality by the axioms, which we claim are "obviously" rational (in some undefined way!)
- Foundation of finance theory.

Assessing Probability I: Relative Frequency

- Where do the numbers come from?
- We can *measure* probability of an event of some type empirically by
  - counting the frequency with which it is observed in a repeated context,
  - comparing it with the frequency of occurance of the context, and
  - *assuming* that this relative frequency is a law that holds whenever the situation arises.
- An example of a "situation" is tossing a coin, and the event is that the coin falls "head up". It's easy to imagine repeating this situation many times to determine if the coin is "fair."

Assessing Probability II: Symmetry

- We can derive probability theoretically from a set of *symmetric* events by giving equal probability to each member of the set.
- A coin is "fair" if it obeys this symmetric law of probability, and therefore the probabilities of heads and tails are each equal to 1/2.
- The *sum of dots* on two dice is not symmetric, but the *pairs of dots* is (if considered correctly, *i.e.*, as *ordered* pairs so that $(1,2) \neq (2,1)$).

Assessing Probability III: Subjectivity

- If we observe a rational actor who chooses to "bet" on event A rather than event B although we know that they are equally *profitable*, we infer that the actor assesses higher *probability* for A than for B.
- We say that such an agent has *subjective probabilities*.
- The agent may have reasons (such as an empirical frequency or a theory based on a symmetry) for the assessment but we don't know them. We often aren't sure the agent has reasons at all.
- If the agent tells us he doesn't care whether to bet on heads or tails with equal payoffs, we know she gives equal probability to each outcome.

Assessments and Mathematics

- In probability theory, we don't care where the assessments come from, we just calculate according to some rules.
- When applying probability to decision theory or to statistical study, we need to make some assumptions about existence and stability of the probabilities across trials.
  - These assumptions are derived from some domain of knowledge. As long as the assumption obey the laws of probability, probability theory can't help choose good assumptions. It can only rule out impossible (inconsistent) assumptions.

Rules of Probability

- The rules of probability are based on *set functions*.
- Events are modeled as sets, which allows definition of rules for combining events logically.
- A *probabilistic model* is a function which gives a numerical value to certain sets, and obeys certain rules for the value of combined sets.

Example: Choosing a Chair

- A committee has 6 members: Ann, Bob, Carol, Don, Ed, and Frank.
- Nobody really wants to be the chairperson, nobody really hates the idea, and they decide to choose by lottery.
- Ann has a 6-sided die, and proposes the rule that each person has a number: Ann $\leftrightarrow$ 1, Bob $\leftrightarrow$ 2, Carol $\leftrightarrow$ 3, Don $\leftrightarrow$ 4, Ed $\leftrightarrow$ 5, and Frank $\leftrightarrow$ 6.

  Rolling the die will generate a number, and the corresponding person becomes chair of the committee.
- We intuitively ("naturally") think that each person can become chairman "1/6 of the time". Although we don't have to choose this as the "probability," it's convenient and symmetric.
- "Ann becoming chair" is the way we describe a certain event in words. As a *set*, we describe it as "$\{\text{Ann}\}$" (not as "Ann"!) Then another event of some interest is "$\{\text{Ann}, \text{Carol}\}$," which we describe in words as "the chair is female." There may be many ways to describe the same event, such as "the chair wears high-heeled shoes".

Introduction to Probability

- Basic probability theory is based on numerical assessments of "likelihood" of *events*.
- Events are treated using *set theory*, and they can be combined and analyzed as sets are.
- Sets can be described explicit lists of elements (one by one), or as collections that satisfy properties. Element lists aren't very useful in probability theory.
- Sets described as satisfying a property connect events to logical propositions.
- We use the rules of logic to combine events, defined by such *properties* of the elements of the sets we think of as events.
- It is useful to consider all possible intersections of events, which give events with "few" elements, or satisfying very precise descriptions. A *partition* is a set of non-empty events such that the union of all of them is the certain event, and no two events in the partition intersect.

Assessing Probability

- Probability obeys some convenient mathematical laws.
- These laws determine the relationship of probability of an event to certain other events, but not the numerical value. There are three general approaches to assessing numerical probabilities:

  Frequency approach The relative proportion of times a specific event occurred in the past is taken as the probability. *Example: weather prediction.*

  Symmetry approach two events which are physically the same are assigned the same probability. *Example: rolling a die.*

  Subjective approach Let an expert estimate it, or use *revealed preference*: if an agent can bet on two events A and B, they have the payoff to the agent, and the agent chooses A, then A has higher *subjective probability* for the agent.

Operations on Sets

$A \cup B$	union	the set of things that are members of at least one of A or B;
$A \cap B$	intersection	the set of things that are members of both A and B
$A \setminus B$	set difference	the set of things that are members of A but not B
$A \subset B$	subset	every element of A is an element of B

Examples, where:

$$\begin{aligned} A &= \{\text{scissors, paper, stone}\} \\ B &= \{\text{needle, thread, scissors}\} \end{aligned}$$

$$\begin{aligned} A \cup B &= \{\text{scissors, paper, stone, needle, thread}\} \\ A \cap B &= \{\text{scissors}\} \\ A \setminus B &= \{\text{paper, stone}\} \end{aligned}$$

Set Operations on Events and Special Events

$A \cup B$	disjunction	the event where the event described by A, the event describe
$A \cap B$	conjunction	the event where both the event described by A, and the event
$A \setminus B$	exclusion	the event where A occurs but B does not
$A \subset B$	implication	if A happens, then B does too
$\emptyset$	null event	"nothing happens", the *impossible* event, or *empty* event (her
Ω	certain event	the disjunction of all possible events, thus "something happen

Probability and Special Events

- Each event has a probability (number) assigned to it. (In mathematics this is called a *set function*.) The probability of an event E is denoted $P[E]$. Some times "P" is spelled "Pr" or "Prob".
- The set of all elements of all events we are interested in is called the *certain event* or the *sure event*, and often denoted by Ω. The probability of Ω is 1.
- The empty set is called the *impossible event* and has probability of 0 (if you add no numbers, you have nothing, *i.e.*, zero). It is denoted by $\{\}$ or $\emptyset$.

Events and Probabilities

- Probabilities are not determined by mathematical theorems: they must be measured, deduced from assumptions, or estimated subjectively. However they do satisfy certain laws:
  - $P[\emptyset] = 0$ and $P[\Omega] = 1$.
  - $0 \leq P[A] \leq 1$ for all events A.
  - If $A \subset B$, then $P[A] \leq P[B]$ for all events A and B.
  - If $A \cap B = \emptyset$, then $P[A \cup B] = P[A] + P[B]$ for all events A and B.
- These are a sufficient set of axioms for countable Ω. There are many other probability laws that can be deduced from these.
- Note that these laws do not refer to elements!

The Sure Event

- It is very common in probability modeling to use uniform probability on some set Ω to structure the probabilies (as we did with the "chair selection problem").
- Usually these are finite sets, the unit interval $[0, 1]$, or products of such sets.
  - *Example 1*: Consider the chair selection problem and define Ω as the set of possible pairs of numbers from a red die and a blue die with equal probability for each pair of numbers. The event "Ann becomes chair" is defined by the set of pairs where the sum of the dice is divisible by 6, "Bob becomes chair" by all pairs among those remaining divisible by 5, and so on for Carol, Don, Ed, and Frank.
  - *Example 2*: Any continuous increasing function $f : [0, 1] \to [0, 1]$ with $f(0) = 0$ and $f(1) = 1$ defines probabilities for all subsets of $[0, 1]$ via the Stieltjes integral (we will call it "cumulative distribution function"). It defines probabilities of intervals by subtracting the values at the endpoints, and (almost) everything else (of interest) by unions and intersections.

Homework

Consider the "chair selection problem" of *Example 1.*

1. What are the probabilities of each pair of *numbers*?
2. What events (as sets of number pairs) correspond to the events that each committee member is selected as chair?
3. Is it possible for all members to become chair? If not, which members are possible, and which impossible?
4. Explain why the definition says "all pairs remaining."
5. What are the probabilities of each member becoming chair?