Various takes on Mathematics and Modelling

B. Mazur

January 8, 2014

Contents

Part I

About our seminar-course "Reasoning via Models."

1 In anticipation

My motivation for participating in this seminar is that it gives me the opportunity to learn about *models*—that ubiquitous tool in aid of "thinking"—from (and in the company with) my two co-teachers, Amartya Sen and Eric Maskin, who have great experience both practical and theoretical, over a vast range. And to learn with—and from—our students, in what I hope will be a fully engaging interchange. In mathematics, *models* play a number of important specific roles, so I'm happy to live—this semester—with the broader concept. I have often found it fruitful to *live with* a concept for a significant length of time—without specifying a particular goal other than to become at home with, intimate with, the concept in broad terms in its various facts and its various moods. Such an experience can provide resonances which enrich thoughts that one may have, or can connect with ideas that one encounters, years later. Everyone in this room has interacted, in one way or another, with the issue of modeling, in one form or another, and this seminar is a way for us to learn about different experiences and viewpoints towards it.

2 What are models?

Models can be: formulations, abstractions, replicas, idealizations, metaphors—and combinations of these. They can be exemplars to be copied (nude models) or to be exemplified (role models).

Generally we will be thinking of a model as a *representation* of some "thing" (which itself could be a network of objects, or ideas) this representation having some focused intention–the intention

being either made explicit, or understood.

The *representation* may represent the "thing" in a setting vastly different from its natural context. For example, the "thing" might be some aspect of the natural world, and the representation might be in the vocabulary of mathematics. If the "thing" has already been described, the model might introduce entirely new vocabulary offering an entirely new description. For example, electric circuits as analogous to—i.e., as modeled by—hydraulics: water flow. A more ubiquitous example is given by the infinitely subtle metaphor, elusively simple, yet understood by absolutely everyone: time as represented by distance. We say, without thinking, "far in the future" or " a long time ago."

A model may be a schematic simplification, a clarification, a projection, a way of distinguishing certain features of the 'thing." More radically, it may be a way of blurring much of the "thing" retaining only one of its aspects. A cross-section, for example. Mathematics abounds with such projections. For example, given the profound and difficult interaction, in number theory, of the infinitely many *prime numbers*, mathematicians have devised a way to project all that number theory to a curious world where there is only *one* prime number, say the prime 2, or 3, or 5, ..., and thereby focus on issues pertaining specifically to that one prime number in a simpler—in vitro, so to speak—'number theoretic' situation which might shed light on the full world of numbers[1].

A model may be a specific change of some parameter, or parameters, a *scale model*, one aim being to retain, or at least faithfully reflect, the full network of interrelations that the initial object would still have, if *scale* were allowed to vary. Another important aim would be to make it small so as to be able to encompass it, all at once, physically.

Extremely useful are "working models," by which I mean structures we can deal with, work on, move its moveable parts, ask precise questions about, so that we might learn about the initial "thing." Often we would be unable to do this—as efficiently—by thinking directly—or by more passively musing— about that thing itself.

A model can precede the "thing itself." It can be a blueprint (to use an old-fashioned word). The oldest vigorous discussion of "models as blueprints" that I'm aware of occurs in Plato's *Timaeus* where the creation of the world is said to be effected by the demiurge according to a *paradigma*, a word we might simply translate as *model*. This fits into the recurring theme in Plato of the role of *image*. Proclus's discussion of this (in his *Commentaries on the Timaeus*) complicates the issue substantially[2].

Models can also have normative missions, and in many of our encounters with them, it is that type of mission—i.e., influencing some practical, or political, change—that is their primary importance.

[1]This 'world' is called the field of p-adic numbers, where p is the particular prime number being examined; e.g. $p = 2$, giving the '2-adics,' $p = 5$ giving the '5-adics,' etc.

[2]Among all the other perplexities that Proclus deals with, he gets into a "dancer from the dance" tangle when he wonders–vis a vis Plato's *Timaeus*, "how comes it to pass that ... the paradigm is one thing, and the Demiurgus another."

3 Comments on the expected format of our seminar

Except for this first ("introductory") session and the final ("wrapping up") session, each of our other sessions will be 'chaired' by one of the professors. So each of us will chair four session. To say that we each 'chair' a session means that we expect full involvement of students in discussions and also, at times, presentations. That is, besides a final paper for the course, we may request a (usually very short) presentation on the part of some volunteers.

For example, in anticipation of each of my four chaired sessions, a student (or a team of students) may be asked (or may volunteer) to give a twenty-minute presentation of specific related material, and to offer handouts related to that material for the class to read—-preferably beforehand.

Here are the specific slants to the subject that my four chaired sessions will take.

1. **General introduction** to different "classical mathematical models" (in somewhat applied contexts). We restrict our attention to the least technically demanding examples.

   (a) *Themes:*

   - A Malthusian discussion.
   - One elementary example, having no differential equations; e.g., *The geometry of blood vessels* (pp.71-73) in [B].
   - A brief mention of a selection of models. E.g., following the Malthus discussion, a bit of Lottka-Volterra differential equations.
   - If time allows, a comprehensive going-through of either the models for biodiversity on Islands following MacArthur and Wilson (but taken from loc. cit.) or perhaps Chapter 2 ("Evolutionary branching in a classical model for sympatric speciation") of [D], *Adaptive Diversification,* Princeton University Press (2011).

   (b) *Possible student presentation:*
   To give a presentation of one of the simpler examples of models in [B] and to explain how the model helps.

   (c) *Reading (to be made more specific soon)*

   - Read in [Malthus] and [R]
   - Browse through [B]
   - Chapter 2 of [D]

2. **Axiomatic methods, their evolution, their models and their applications.**

   (a) *Themes:*

   - The 'common notions' and discursive structure of mathematical arguments in Euclid.
   - Peano axioms– further re-interpreted by Hilbert in terms of his 'formal systems'.

- Formal systems, as a mathematical object. A model for a formal system implying its *consistency*.
- A discussion of the difference between a formal system allowing utterly different models, and one that is meant to be have, speaking informally, a categorical model.
- A bit of Gödel. (E.g., *Gödel numbering* as a model.)
- Fifth postulate issues, non-Euclidean geometries.
- The hyperbolic plane in its various models.
- More Gödel.
- Algorithms.

(b) *Possible student presentation:*

- A discussion of the definition of formal system
- The history of the Fifth postulate issues
- The hyperbolic plane in its various models.

(c) *Reading (to be made more specific soon)*

- Short selections in [H].
- Short selections in [G].
- Short excerpt from Hilbert's essay "On the Infinite" (in [H]) but specifically: [H1].

3. **Probabilistic models**

(a) *Themes:*

- The heuristic of *reasoning from randomness* (e.g. [M1]).
- Bayesian formulations. For this I have a handout [M2] and also some selections from Nate Silver's [S].
- Monte Carlo methods, maybe (and maybe from pp. 103-120 of [B]).

(b) *Possible student presentation:*

- An example of application of the Monte Carlo method.
- Modeling prediction biases. As discussed in various chapters of [S] (e.g., pp.197-203.
- The 'discussion' between Peter Norvig and Noam Chomsky regarding "the two cultures of statistical learning" in [N].

4. **Axiomatic methods II.** This session will be devoted to prior topics that overflowed their time limit, and therefore were not fully covered previously. But depending on time available, and the nature of our audience, we can go further in one (or both) of two directions:

- We could devote as much of the session as available to a general discussion of the current curious debate about *small structural models versus 'Big Data'* (as in the Norvig/ Chomsky material [N]), or

- we can do a bit more technical mathematics, dealing with what one might call *Axiomatic methods II:* Categories and Functors that allow one mathematical theory to model another. And at least mention Model Theory (and the Lowenheim-Skolem theorem).

References

[B] E.A. Bender, *An introduction to Mathematical Modeling,* Dover (2000)

[D] M. Doebeli, *Adaptive Diversification,* Princeton University Press (2011)

[G] M. Greenberg, *Euclidean and non-Euclidean geometries: Development and History* (3rd Edition) Freeman (1993)

[H] J. van Heijenoort, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931* (Source Books in History of Sciences)

[H1] D. Hilbert, Excerpt of "On the infinite." This is the attachment "Hilbert.Excerpt.pdf." The entire essay is in [H].

[Malthus] Thomas Malthus, *An Essay on the Principle of Population* (1798) Text: `http://www.esp.org/books/malthus/population/malthus.pdf`

[M1] B. Mazur, Is it Plausible? The Mathematical Intelligencer (onLine) August 2013

[M2] B. Mazur, Data and Hypotheses (this is the attachment "bayes3.pdf.")

[N] P. Norvig, *Colorless green ideas learn furiously; Chomsky and the two cultures of statistical learning* pp. 30-33 in *Significance* The Royal Statistical Society, August 2012. (This is in the attachment:"Norvig.Significance.pdf")

[R] E. Rothschild, *"Axiom, theorem, corollary &c.": Condorcet and mathematical economics,* pp. 287-302 in Soc. Choice Welfare **25** (2005)

[S] N. Silver, *The Signal and the Noise,* Penguin Press (2012)

Part II

Notes for the session: *An introduction to mathematical models*

4 'Intention' and modeling

In our introductory session we pointed out that a *model* is a representation of some "thing" (which itself could be a network of objects, or ideas) and will have some focused intention–the intention being either made explicit, or understood.

Models might have normative missions. Or their goal might be to explain some specific phenomenon. Often this intended function—i.e., influencing some practical, or political, change, or explaining a very specific aspect of a perplexed problem—is the primary importance of the model. Eric Maskin's illuminating presentation of *models as parable* highlighted such intentions: parables, after all, have morals[3].

Given this official sense of the term 'model,' some of our *mathematical models* will be more like templates for models—i.e., allowing a broad variety of intentions—than fully-fledged intention-driven models.

But, of course, nothing is intention-free. Even one of the basic sources of vocabulary for mathematical models—i.e., Statistics— hides in its name a normative role. To quote D'Arcy Thompson,

> For Shakespeare, or for Milton a *statist* meant (so Dr. Johnson says) "a politician, a statesman, one skilled in government. The eighteenth century *Statistical Account of Scotland* was a description of its State...

and even earlier theoretical categories that carry their intention in their very name include *Political Arithmetic* and *Physique Sociale*.

So it is no surprise that Malthus's treatise isn't a pure contemplation of population variation. Views regarding human perfectibility and a controversy (related to Condorcet) feature in it. But we'll focus on the mathematical model per se.

5 Malthus

Thomas Malthus (*An Essay on the Principle of Population 1798*) gives himself two starting postulates:

> First, that food is necessary to the existence of man.
> Secondly, that the passion between the sexes is necessary and will remain nearly in its present state.

What gets Malthus going is the disparity of rate of increase of the first necessity, food, as compared with the rate of increase of population, given the second postulate.

"Population, when unchecked," writes Malthus,

> increases in a geometrical ratio. Subsistence increases only in an arithmetical ratio. A slight acquaintance with numbers will shew the immensity of the first power in comparison of the second.

[3] If time permits, I'll make some comments about the formal shape of some of the parable models we saw last time.

The main object to 'measure' is a *population*—counting the number of distinct individuals—and more specifically to understand how this count changes in time. Malthus's *geometrical ratio* hypothesis is essentially the assumption is that the *rate of change of the population* at any moment in time, t, is proportional to the size of that population at that moment. This, of course, is but a starting assumption, to be modified with the complexity of the model, which—to be in any serious way realistic— will certainly have to bring in limited food supply as Malthus does, but may bring in other actors (e.g., predators) and other constraints (e.g., disease) as well.

But the naive assumption unadorned with any baroque complications (that the rate of change of the size of the population is just proportional to size of the population) would predict *exponential growth*–e.g., twice the population twice the rate of growth, etc. Surely an unsustainable state of being, but simple enough. It is traditional to put it in the form of a differential equation. So, letting $P(t)$ be the size of the population at time t

$$(*) \qquad \frac{dP}{dt}(t) = \text{ some constant} \times P(t).$$

This equation is indeed nothing more than a direct translation of the statement 'rate of change of population is proportional to size of population.' Since we haven't specified what units we use to parametrize time t, and since we haven't specified the units that describe P as well in cases where the 'size of population' is not given as a discrete number of individuals, but as a somewhat continuous quantity, we haven't really specified the 'some constant' in this equation which will depend on these choices. Moreover, if P is, in fact, a 'number of individuals' this equation can be nothing more than shorthand for a Difference Equation approximation to it[4].

That being said, the solution to the differential equation **(*)** can be expressed as:

$$(**) \qquad P(T) = P(0) \cdot 2^T$$

where $P(0)$ is the population size at time 0(= "NOW') and T is time proportional to the t of the initial equation **(*)** and measured in terms of 'doubling time units' for the specific population under study (and this is not necessarily well-known units like hours or minutes or seconds). In each unit of doubling-time—in each tick of the T-clock ($T \mapsto T + 1$)—this population will double (according to this model)[5].

One way of viewing this Malthusian equation is as not yet a full-fledged *model* but as an 'opening move' in what will be a possibly never-finalized attempt to construct a model that reflects all the specific understandings that one achieves as one studies more deeply whatever situation it is that one aims to 'model." An armature, if you wish.

Malthus, himself, takes that approach in his essay. For starters, he emphasizes that limitations of the rate of expansion of food supply goes directly counter to any expectation that in the long

[4]We will see more about this distinction between *differential equations* and *difference equations* later, but for now let me say that with the exception of the very few differential equations that admit simple exact solutions (of which this is one!) all the others are indistinguishable from difference equations, for that is how computers interpret them in order to determine their numerical solutions.

[5]Malthus estimated that North America—at the time of his writing—was 'doubling' its population every 25 years.

term, the ratio dP/dt over P (i.e., rate of change of population compared to size of population) be constant. This gives at least one clear mechanism of *population self-regulation* in the terminology, of some modern writers; see the interesting discussion-article by Peter Turchin, *Does population ecology have general laws?* OIKOS **94** 1726. Copenhagen (2001).

Of course, once you decide that the ratio $r := \frac{1}{P}dP/dt$– i.e., the *rate, per person, of population increase or decrease*—needn't be constant, and therefore you've opened up the question of its variation (so, $r = r(f_1, f_2, \ldots, f_\nu; t)$ can change as a function of time t, and possibly other parameters $f_1, f_2, \ldots, f_\nu$) in a sense you've kicked the model down the road, because you have no longer made any assertion at all yet about the behavior of population size; all you have done is to have framed an open-ended question: how, and depending on what, does r vary?

One common modification of the equation to accommodate for the expectation that any fixed environment will only support a maximum population[6] (call it M) is often called the logistical equation. Namely, you assume that the environment has indeed a saturation point and therefore you seek a "parable-model," so to speak, that reflects this. A good exercise is to think, before looking at the equation below, how you would fashion a relatively simple modification of the Malthusian equation that expresses this saturation. OK, here is one favorite:

$$r(t) = \frac{1}{P(t)} \cdot \frac{dP}{dt}(t) \;=\; k \cdot (M \;-\; P(t)).$$

or, by moving terms around,

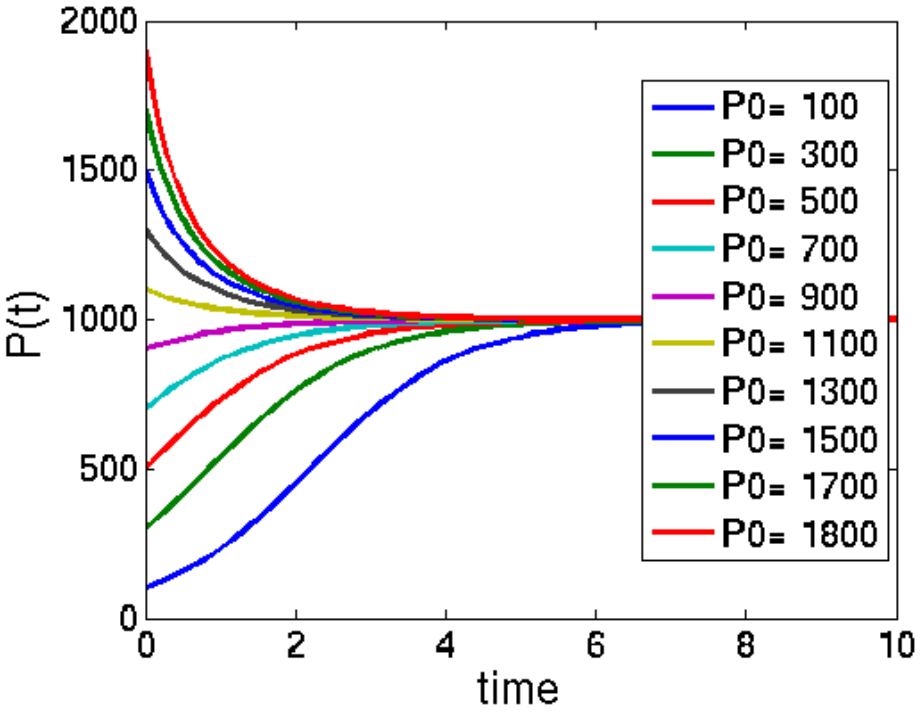
$$\frac{dP}{dt}(t) \;=\; kM \cdot P(t)(1 \;-\; \frac{P(t)}{M}).$$

Here is a graph[7] of a solution with $M = 1000, k = 10^{-3}$, i.e., for the differential equation:

$$\frac{dP}{dt}(t) \;=\; P(t)(1 \;-\; \frac{P(t)}{1000}).$$

[6]For a curious counter-claim published in an Op-Ed essay a few days ago in the New York Times, see *Overpopulation Is Not the Problem* by Erle C. Ellis, begins with the proclamation that "There really is no such thing as a human carrying capacity on the earth," and is written as "the antidote to the demographer and economist Thomas Malthus and his theory that population growth tends to outrun the food supply." `http://www.nytimes.com/2013/09/14/opinion/overpopulation-is-not-the-problem.html?emc=eta1&_r=0`.

[7]taken from the web: `http://images.lingvistika.org/w/images/b/b2/Hoppensteadt_pp.gif`



Malthus does push this discussion further to describe—among other things—an intrinsic "oscillation," as he calls it, commenting that it "will not be remarked by superficial observers." He writes:

> We will suppose the means of subsistence in any country just equal to the easy support of its inhabitants. The constant effort towards population, which is found to act even in the most vicious societies, increase the number of people before the means of subsistence are increased. The food therefore which before supported seven millions must now be divided among seven million and a half, or eight million. The poor consequently must live much worse, and many of them be reduced to severe distress. The number of labourers also being above the proportion of the work of the market, the price of labour must tend toward a decrease, while the price of provisions would at the same time tend to rise. The labourer therefore must work harder to earn the same as he did before. During this season of distress, the discouragements to marriage, and the difficulty of rearing a family are so great that population is at a stand. In the mean time the cheapness of labour, the plenty of labourers, and the necessity of an increased industry amongst them, encourage cultivators to employ more labour upon their land, ... till ultimately the means of subsistence become in the same proportion to the population as the period from which we set out. The situation of the labourer being then again tolerably comfortable, the restraints to population are in in some degree loosened and the same retrograde and progressive movements with respect to happiness are repeated.

To put Malthus's idea of oscillation into some kind of mathematical vocabulary[8] let us give ourselves the letters F for food supply, C for cost of provisions, L for number of labourers, W for wages, i.e., price of a labour-hour, and r, as above, for rate of change of population. So, Malthus argues: as P goes up, F goes down essentially linearly per capita, so C goes up causing distress which makes r go down. But even a small exponential is an exponential, so L goes up even though the amount of work necessary doesn't require such a high L causing W to go down, so r goes further down, so P catches up with supplies of provisions forcing a reversal of all the tendencies listed.

If we imagine this made more precise, say with appropriate time-lags and guesses for the general shape of all these dependences, we would be looking at a finite system of linked Difference Equations that would animate this Malthusian oscillation.

A more vigorous discussion of rates of population increase or decrease would acknowledge that there is a spread, in the population, of different ages, each with its expected mortality statistics, and how that age-group contributes or doesn't, to population increase. There might well also be some skewing of the numbers if significantly more than half, or less than half, of the population were women. So, what if we imagine n different age groups—each spanning a single decade, say— going from youngest to eldest and we only tabulate population of women in these groups $P_0(t), P_1(t), \ldots, P_{n-1}(t)$ as functions of time (1 unit = 1 decade). Each group ($i = 0, 1, 2, \dot{;} n-1$) will have a typical "birth rate" $0 \leq b_i$ (giving the rate of birth of girls) and a "survival rate" $0 \leq s_i \leq 1$. So, that, for example, the population registered as $P_0(t)$ is pre-puberty and therefore will presumably have $b_0 = 0$. Also, let's round off the mortality statistics to $s_0 = 1$, while the post-menopausal decades will have $b_i = 0$. we then might reason that

$$P_0(t+1) \ = \sum_{i=0}^{n} b_i \cdot P_i(t)$$

while

$$P_{i+1}(t+1) = s_i \cdot P_i(t)$$

for $i = 0, 1, \ldots n-1$. Or arraying the population data for each decade as a column vector

$$\mathcal{P}(t) := \begin{bmatrix} P_0(t) \\ P_1(t) \\ P_2(t) \\ \ldots \\ P_n(t) \end{bmatrix}$$

$$\mathcal{M} := \begin{bmatrix} b_0 & b_1 & b_2 & \ldots & b_n \\ s_0 & 0 & 0 & \ldots & 0 \\ 0 & s_1 & 0 & \ldots & 0 \\ 0 & 0 & s_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & s_n \end{bmatrix}$$

So we have the recurrence relation,

[8] Vaguely analogous to this is the type of oscillation occurring in the solutions to the standard predator-prey (Lotka-Volterra) differential equations.

$$\mathcal{P}(t+1) = \mathcal{M} \cdot \mathcal{P}(t)$$

There is quite a literature on issue of stability and near stability regarding such recurrence relations, and one name attached to this is Alfred Lotka[9].

But for a simply marvelous account of all this material and more, very richly and yet succinctly done, read *The growth of a population*, pp. 142-159 in D'Arcy Thompson's *Growth and Form*, Cambridge University Press (1948).

Returning to Malthus's own discussion in the quoted paragraph, we could note that it already provides us with a model featuring a certain type of interlinked dependencies. The Difference Equations I had alluded to—but hadn't, in fact, written down—are not meant to constitute an independent model, but rather to be a faithful translation of Malthus's discussion into mathematical terms, and therefore should be subservient to Malthus's description. If those equations mystify, they will have failed their mission. Moreover, it is legitimate to ask, without prejudice, what–if anything–is the 'value added' in the act of mathematicizing anything? In particular why might we want to provide a mathematical formulation of the paragraph I quoted from Malthus's essay? Here is a tentative list of reasons.

- *Succinctness, possibly:* you would have a linked set of equations encapsulating Malthus's discursive description. A handy mnemonic, even if nothing else.
- *Quantification:* you would be forced to define the variables explicitly, as measurable quantities.
- *The equations can serve as a receptacle:* you might not yet know, or yet want to specify explicitly, the various dependencies listed above, but you might rather wish to allow for some–even if not infinite—flexibility: e.g., r might depend on C and W and even on P, but you might need more data or more experience before you stipulate anything precise about that function $r(C, W, P)$. Even with this type of 'blanks to be filled in later'–e.g., what explicitly is this $r(C, W, P)$?—these equations might well provide a working vocabulary on which to pin whatever you later learn, the dependencies to be specified ever more precisely as time goes on.

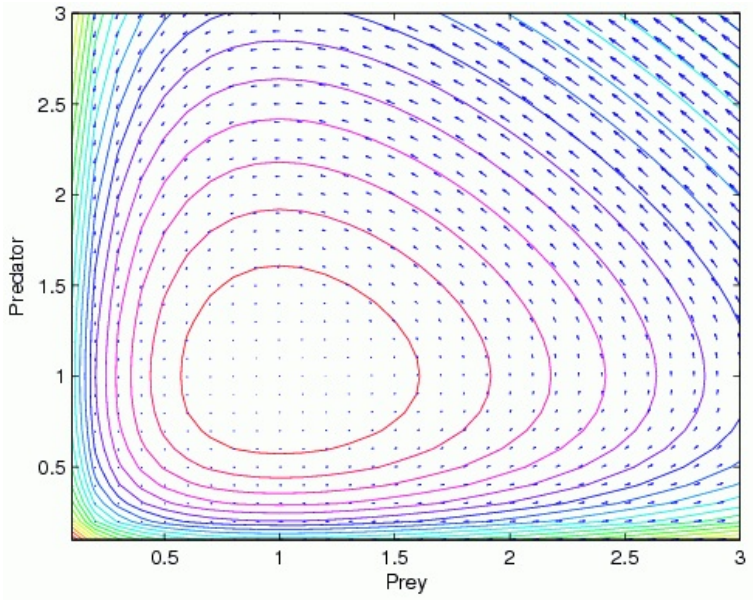
  One opportunity (or perhaps danger) here–once one sets about making guesses regarding the relationship between distinct variables—is the irresistible urge to make exceedingly precise (perhaps over-precise) guesses about these relationships. This is a freedom that may allow interesting experimentation, or—motivated by convenience—may provide a simplicity to the model permitting actual close study. *Creative over-precision,* to put a good face on it. This may be very instructive and a good thing to do, and not at all akin to the mischief of over-precisely calculating quantities to ten decimal places when the margin of error of the calculation would make most of those decimal places meaningless.

[9]For brief nontechnical sketch of some of Lotka's interests, see Jacques Véron's *Alfred Lotka and the Mathematics of Population* in the Electronic Journal for History of Probability and Statistics, **4** (2008) `http://www.jehps.net/juin2008/Veron.pdf`.

Here's an example. The totally general—and therefore not yet usable—predator-prey pair of differential equations has the form

$$\frac{1}{P}\frac{dP}{dt} = f(P,Q); \qquad \frac{1}{Q}\frac{dQ}{dt} = g(P,Q)$$

where P and Q are the population sizes of two species that interact with each other, affecting each other's population growth rate (predator and prey being the eponymous example) and where the functions f and g—as yet unstipulated–describe how these rates are affected. Useless, so far if we are given no hint about what the relations f and g are meant to be. Now the simplest, if not the most realistic guess you might make about these functions is that they be linear functions of P and Q. Here is is a picture[10] of the cycles one gets when one puts such linear functions $f(P,Q)$ $g(P,Q)$ on the right hand side of the equations displayed above:



If you go simpler than linear functions

$$f(P,Q) \quad := aP + bQ + c \tag{1}$$
$$g(P,Q) \quad := cP + dQ + e, \tag{2}$$

you hardly have posited any interaction at all between your two species. To be sure you would naturally choose appropriate signs for the coefficients of your linear functions in order to model relationships that might be labeled predator/prey. You end up with the classical Lotka-Volterra equations. Are they a realistic model for any actual predator/prey interaction? I wouldn't know, but I already am happy that–Occam's razor-style–this system of differential equations stands a chance of being, perhaps, the right start of a discussion; it is an interesting, simple, manageable, toy mathematical machine that might, after being subjected to appropriate modifications dictated by actual experience, approximately model some real-world (predator versus prey) behavior. Here, though, the warning applies: if you are a firm aficionado of these elegant differential equations it is deeply rewarding to study

[10] taken from the web: `http://images.lingvistika.org/w/images/b/b2/Hoppensteadt_pp.gif`

them and their solutions with whatever exactitude you can achieve. Which is a wonderful thing to do in itself. But if you aim to be building some model, unless you also have some control over the stability of the qualitative aspects of your solutions—stability under minor modifications of this idealized model–you run the risk of being a tad too over-precise in your analysis if you dote on the elegant specifics of its solutions. For more on these equations, see: `http://www.scholarpedia.org/article/Predator-prey_model`.

- *Numerical experimentation:* Once the equations become specific enough you can run computer experiments allowing you to visualize the concrete effect of these interlinked dependencies.
- *Surprise or Confirmation:* When you run these equations numerically, you might be surprised by the outcome, or find that your qualitative expectations are confirmed. But, with any such surprise you certainly can, and possibly should, raise the question: does this surprise point to something legitimate, or is it a warning-signal that my mathematical translation was flawed?
- *The 'next question':* You might be led to ask questions on a finer level.

6 Arguments of Scale

A spherical object has its surface area proportional to the square of its radius, and the two-thirds power of its volume. But how long can a dachshund be?

Our text, pp. 26-29 of Bender discusses the issue of *deflection* δ of a beam of length ℓ, vertical thickness t, cross-sectional area A, and subject to a uniform load F with endpoints secured. The model that is discussed takes its input from the theory of elasticity which tells us that:

$$\delta \propto \frac{F \cdot \ell^3}{t^2 A}.$$

Once we are given that formula, it is easy enough to work with, but the point I want to bring out is how natural it is to make reasonable simplifying assumptions that help us work effecgtively with equations given by the physics of the matter.

This bears, a bit, on the Hillary Putnam essay which was on the reading list for last session,. There, Putnam makes a distinction between two entities in any application of a science such as physics to any actual physical event: there are the physical (universal) laws themselves, but also the battery of *auxiliary statements* that make, in effect, powerful empirical judgments regarding how to condition, or to interpret those laws.

Here we assume that F is proportional to $\ell \cdot A$ which is fair enough, that ℓ^2 is proportional to A (thinking of animals with cylindrical trunks) but we also judge that the maximal deflection tolerable would be directly proportional to the length, i.e., one works up a model with the constraint that $\frac{\delta}{\ell}$ is constant. One ends up with the conclusion that

$$t \propto \ell^{3/2}$$

or: the 'prediction' that studying a range of animals of different sized you won't simply find that the larger ones tend to have the shape obtained by zooming the dimensions of the larger ones. Rather, that 3/2 in the exponent suggest that longer animals will have significantly thicker trunks (which seems to be the case).

Bender refers to books by (J.B.S.) Haldane, Rashevsky, and Maynard-Smith for further discussion of scale. But to my mind, the most beautiful account that gets to the heart of the issue rapidly is (again) D'Arcy Thompson (pp. 22-40 of Chapter II (On Magnitude) in *Growth and Form.* There he moves from the sketch of the analysis we just did to a discussion of the diameters of trees, and then to *Froude's Law* which gives that the velocity of a fish or a ship is proportional to the square root of its length. Here is a challenge: can you work up a model, making bold but reasonable assumptions, that yields Froude's Law? (Note, though, that we are assuming that the general proportions of the type of fish or ship we are considering is fixed, except for scale.)

Galileo (1638) opens his *Dialogues Concerning Two New Sciences* with a discussion of the issues of scale and on the "second day" (Salviati) gives an account of the exponent 3/2, as discussed above[11]: (Note below that *sesquialteral* means in the relationship of 3 to 2.)

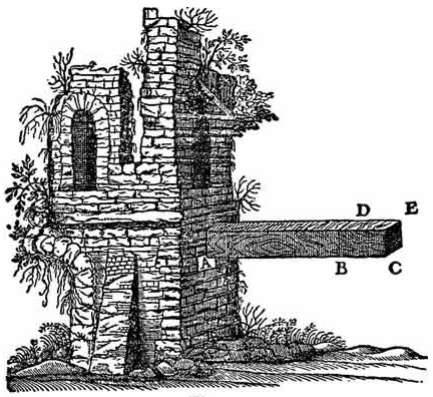


Fig. 19

[11]For a hint that discussions of beam-deflection issues can already be found in the pseudo-Aristotelian literature, consider the following problem posed in the text *The Mechanical Problems of Archytas of Terentum:*

> **Problem 16**
>
> Why is it that the longer a board is, the weaker it gets? and, lifted, bends more, even if the short one—say, two cubits—is thin, and the long one—say, 100 cubits—is thick? Because in the lifting, the length becomes lever, weight, and fulcrum? The part in the hand practically becomes a fulcrum, the part at the end becomes the weight, so that the further it is from the fulcrum, the farther it must bend [repeated]. As it is necessary to raise the ends of the lever, so if the lever be bent, it has to bend more on being lifted, which happens with longer boards. With shorter ones, the end is near the unmoving fulcrum.

Proposition VI

In the case of similar cylinders and prisms, the moments [stretching forces] which result from multiplying together their weight and length [i. e., from the moments produced by their own weight and length], which latter acts as a lever-arm, bear to each other a ratio which is the *sesquialteral* of the ratio between the resistances of their bases.

In order to prove this let us indicate the two similar cylinders by AB and CD: then the magnitude of the force [momento] in the cylinder AB, opposing the resistance of its base B, bears to the magnitude [momento] of the force at CD, opposing the resistance of its base D, a ratio which is the sesquialteral of the ratio between the resistance of the base B and the resistance of the base D. And since the solids AB and CD, are effective in opposing the resistances of their bases B and D, in proportion to their weights and to the mechanical advantages [forze] of their lever arms respectively, and since the advantage [forza] of the lever arm AB is equal to the advantage [forza] of the lever arm CD (this is true because in virtue of the similarity of the cylinders the length AB is to the radius of the base B as the length CD is to the radius of the base D), it follows that the total force [momento] of the cylinder AB is to the total force [momento] of the cylinder CD as the weight alone of the cylinder AB is to the weight alone of the cylinder CD, that is, as the volume of the cylinder AB [listesso cilindro AB] is to the volume CD [allistesso CD]: but these are as the cubes of the diameters of their bases B and D; and the resistances of the bases, being to each other as their areas, are to each other consequently as the squares of their diameters. Therefore the forces [momenti] of the cylinders are to each other in the sesquialteral ratio of the resistance of their bases.

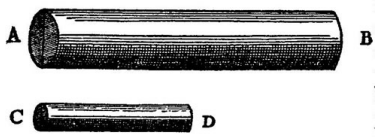


Fig. 23

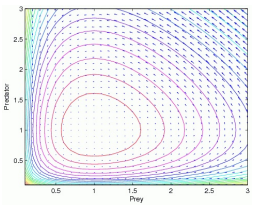
7 More notes for the session: "Introduction to mathematical models"

1. **Predator/Quarry:** Recall the totally general—and therefore not yet usable—predator-prey pair of differential equations which has the form

$$\frac{1}{P}\frac{dP}{dt} = f(P,Q); \qquad \frac{1}{Q}\frac{dQ}{dt} = g(P,Q)$$

where P and Q are the population sizes of two species that interact with each other, affecting each other's population growth rate (predator and prey being the eponymous example) and where the functions f and g—as yet unstipulated–describe how these rates are affected. Here, again, is a picture[12] of the cycles one gets when one puts very simple predator-prey-type functions $f(P,Q)$ $g(P,Q)$ on the right hand side of the equations displayed above:

[12]taken from the web: `http://images.lingvistika.org/w/images/b/b2/Hoppensteadt_pp.gif`



But it might be instructive to (as an exercise) to go more slowly step-by-step in developing various relations between Predator and Quarry. Below, the early lower case letters of the alphabet designate positive constants. The simplest and most ridiculous being:

- **Step 1:**

$$\frac{dP}{dt} = aP \tag{3}$$
$$\frac{dQ}{dt} = bQ, \tag{4}$$

with a and b constants. Or:

Here you have posited no interaction at all between your two species, and their rates of population change would be dictated by two straight independent Malthusian equations $P(t) = P(0)e^{at}$ and $Q(t) = Q(0)e^{bt}$. Or, viewing

$$\mathcal{P} := \begin{bmatrix} P \\ Q \end{bmatrix}$$

as a column vector, and

$$\mathcal{M} := \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

we get the differential equation:

$$\frac{d\mathcal{P}}{dt} = \mathcal{M} \cdot \mathcal{P},$$

leading us

- **Step 2:** to consider this same format for a differential equation coming from more general 2×2 matrices

$$\mathcal{M} := \begin{bmatrix} a & -c \\ -d & b \end{bmatrix}$$

where the off-diagonal terms actually show the two populations engaging. E.g.,

$$\frac{dP}{dt} = aP - cQ \tag{5}$$
$$\frac{dQ}{dt} = bQ - dP, \tag{6}$$

For example, if $c = 0$ we still have the predator–oblivious to the size of its quarry—expanding at the rate $P(t) = P(0)e^{at}$ but the poor quarry subject to the differential equation:

$$\frac{dQ}{dt}(t) = (b - dP(0)e^{at})Q(t),$$

a differential equation in Q (albeit with nonconstant coefficients) whose solutions are plummeting.

Of course, one might be interested in matrices $\mathcal{M}$ that actually lead to *stable* (even if unrealistic) situations, but: for any such "model" linearity implies, as Tom Lovering suggested, doubling the populations P and Q would lead to the same qualitative outcomes. This, of course, is not realistic except under the situation where you you simply have two identical aquaria, so to speak, rather than one, and the aquaria don't connect with one another. Once you do the doubling of populations and these double-populations interact as a twice-large community, you might well expect—for various reasons–that the cross-interactions are non-linear, as they are with Lottka-Voterra. But it is instructive to start adding very small nonlinear contributions to the basic linear differential equation, working up a possible 'model scenario' that explains the appearance of these terms, and numerically solve. Or, one might start with independent "logistical equations" (as in last week's note) and do the same.

2. I want to thank Gabriel Erion of our class for suggesting a few models to think about (one of which has a curious predator/prey feel to it without there being a predator or a prey. It has to do with drops of liquid on a surface causing ripple waves on the surface, and then bouncing ("walking") along the liquid surface as influenced by their own wave.

   These are files:

   Macroscopic Diffraction and Interference.pdf and

   Quantization of Classical Orbits.pdf which I'll put on our web-page.

   Here's a quote from one of these file:

   > We called "walker" the moving droplet dressed with the wave-packet it emits. A walker is a 'symbiotic structure: if the droplet disappears (by coalescence with the substrate), the wave vanishes. In reverse, if the wave is damped, the droplet stops moving. Two interacting walkers can have discrete stable orbits [3,4] demonstrating that they have both an inertia due to their mass and nonlocal interactions due to the interference of their waves.

Part III

Skeletal notes for the session: *Geometry*

8 Definitions in *Euclid's Elements Book I*

1. A point is that which has no part.
2. A line is breadthless length.
3. The extremities of a line are points.
4. A straight line is a line which lies evenly with the points on itself.
5. A surface is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A plane surface is a surface which lies evenly with the straight lines on itself.
8. (Def'n. 13) A boundary is that which is an extremity of anything.
9. (Def'n. 14) A figure is that which is contained by any boundary or boundaries.
10. (Def'n. 15) A circle is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another;
11. (Def'n. 16) And the point is called the centre of the circle.

9 'Postulates' in *Euclid's Elements Book I*

1. To draw a straight line from any point to any point.
2. To produce a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance.
4. That all right angles are equal to one another.
5. *('Fifth Postulate':)* That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

10 Common Notions in *Euclid's Elements Book I*

1. Things which are equal to the same thing are also equal to one another.
2. If equals be added to equals, the wholes are equal.
3. If equals be subtracted from equals, the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

11 Comments on Euclid

On the nature of the definitions, postulates, etc.:

- There are actual arguments, and the format of the arguments is explicit, the parts articulated. This seems, at least by report (via Simplicius) to be the case with even the earliest Greek mathematics that we know (Hippocrates of Chios).
- Euclid is describing *one structure*: the Euclidean plane. This structure (presumably) precedes the discursive analysis: the format of Euclid's axioms hardly accommodates the concept of a 'model of itself.'
- The issue, specifically, of *uniqueness* doesn't enter. E.g.:
  - What about the following 'familiar' definitions, not very much in the spirit of the definitions in Book I?

    A straight line is [implied: uniquely] *determined by two of its points.*

    Or:

    A straight line segment is the [implied: unique] *curve of shortest distance between its endpoints.*

  - Or uniqueness, for example, of that *central point* in Def'ns 15 and 16?

- For me, the most striking fact about these definitions is that they don't rely on set theoretic vocabulary. We moderns immediately think 'sets,' 'subsets,' 'membership in sets,' etc.
- No vocabulary for 'continuous motion,' 'transformation,' 'function' except as these issues are introduced in the *Postulates* and/or when one triangle is "applied" to another. Discuss the 'Erlangen Program.'
- Discuss length and proportion as in the file 'Comments on length'.

On the Postulates:

- Construction rather than Existence
- Fifth Postulate: the minute one questions its independence, one is on the way to model-formation.

On the Common Notions:

These are closest to modern axiomatics.

Hilbert's Euclidean Geometry

Hilbert's axiom system is constructed with six **primitive notions:**

1. three primitive terms:
   - point;
   - line;
   - plane;

   and

2. three primitive relations:
   - *Betweenness,* a ternary relation linking points;
   - *Lies on (Containment)*, three binary relations, one linking points and straight lines, one linking points and planes, and one linking straight lines and planes;
   - *Congruence*, two binary relations, one linking line segments and one linking angles. Note that line segments, angles, and triangles may each be defined in terms of points and straight lines, using the relations of betweenness and containment. All points, straight lines, and planes in the following axioms are distinct unless otherwise stated.

   And there are these structures and axioms:

3. **Incidence** For every two points A and B there exists a line a that contains them both...

4. **Order** If a point B lies between points A and C, B is also between C and A, and there exists a line containing the distinct points A,B,C... If A and C are two points of a line, then there exists at least one point B lying between A and C. Of any three points situated on a line, there is no more than one which lies between the other two.

Pasch's Axiom: Let A, B, C be three points not lying in the same line and let L be a line lying in the plane ABC and not passing through any of the points A, B, C. Then, if the line L passes through a point of the segment AB, it will also pass through either a point of the segment BC or a point of the segment AC.

5. **Congruence** If A, B are two points on a line L, and if A' is a point upon the same or another line L', then, upon a given side of A' on the straight line L', we can always find a point B' so that the segment AB is congruent to the segment $A'B\ldots$

6. **Continuity**

   - *Axiom of Archimedes:* If AB and CD are any segments then there exists a number n such that n segments CD constructed contiguously from A, along the ray from A through B, will pass beyond the point B.

   - *Axiom of line completeness:*....

12 Comments

- An articulation different from Euclid's: the triple *definitions/postulates/common notions* being replaced by *primitive terms/primitive relations/structures and axioms.*
- The common notions (i.e., logical pre-structures like 'equality') are implicitly assumed rather than formulated.
- Modern quantification explicit. E.g., the 'incidence Axiom ' calls up universal and existential quantification: $\forall$ points A, B, $\exists$ a line through A and B.
- Set theoretic vocabulary. Discuss.
- Geometry as a structure.

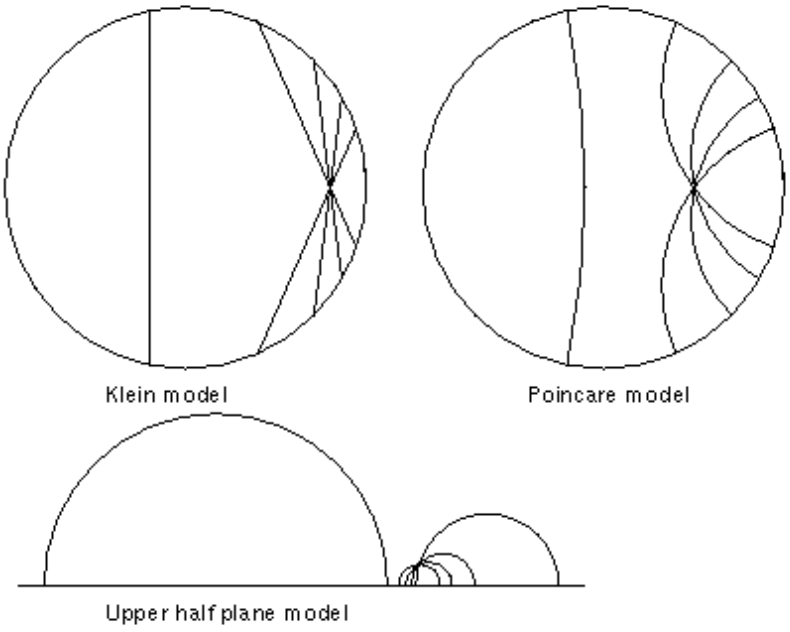
From Hilbert's *On The Infinite:*

> ...We encounter a completely different and quite unique conception of the notion of infinity in the important and fruitful method of ideal elements. The method of ideal elements is used even in elementary plane geometry. The points and straight lines of the plane originally are real, actually existent objects. One of the axioms that hold for them is the axiom of connection: one and only one straight line passes through two points. It follows from this axiom that two straight lines intersect at most at one point. There is no theorem that two straight lines always intersect at some point, however, for the two straight lines might well be parallel. Still we know that by introducing ideal elements, viz., infinitely long lines and points at infinity, we can make the theorem that two straight lines always intersect at one and only one point come out universally true. These ideal "infinite" elements have

the advantage of making the system of connection laws as simple and perspicuous as possible. Moreover, because of the symmetry between a point and a straight line, there results the very fruitful principle of duality for geometry....

Hyperbolic Geometry:

- Klein model (used by many)
- Poincaré disc model (loved by geometers)
- Poincaré half-plane model (loved by number theorists)
- Lorentz model or hyperboloid model (loved by physicists)



A good discussion of this is in `en.wikipedia.org/wiki/Hyperbolic_geometry`

Arithmetic

1. **Propositional calculus— or "zeroth-order logic"**

   - a set of primitive symbols A

     (atomic formulae, placeholders, proposition letters, or variables),

   - a set of operator symbols Ω

     (logical operators or logical connectives such as $\neg, \wedge, \vee, \rightarrow$),

   - "rules" Z

     (such as *modus ponens*),

   - Axioms I

     (such as:

     $$F \rightarrow (G \rightarrow F)$$

     or:

     $$F \rightarrow F \wedge G,$$

     etc.)

   By definition A **well-formed formula** (wff) is any atomic formula, or any formula that can be built up from atomic formulas by means of operator symbols according to the rules of the grammar.

2. **First order logic**

While propositional logic deals with simple declarative propositions, first-order logic also has predicates and some simple type of quantification. A *predicate* I think of naively as hovering between being *intrinsic* (e.g., a "property") or *extrinsic* (e.g., "the set of elements having that property"). In first-order theories, predicates are often simply associated with sets. But we can't freely quantify—in a first-order theory–over sets. This distinguishes 'first-order logic from higher-order logic in which predicates can have predicates or functions as arguments, or in which one or both of predicate quantifiers or function quantifiers are permitted: In interpreted higher-order theories, predicates may be interpreted as sets of sets.

There are two key parts of first order logic.

- *Syntax* The syntax determines which collections of *symbols* are legal expressions in first-order logic, while the semantics determine the meanings behind these expressions. The symbols are often regarded simply as letters and punctuation symbols. But we can divide the symbols of the alphabet into *logical symbols,* which always have the same meaning, and *non-logical symbols,* whose meaning varies by interpretation. E.g., the logical symbol $\wedge$ always represents "or."

- *Alphabet* This consists of:
  - *terms,* which intuitively represent objects, and
  - *formulas,* which intuitively express predicates that can be true or false.

  The terms and formulas of first-order logic are strings of *symbols.*

3. Naive Model Theory

A **model** of a first-order theory ascribes a (so-called: "meaning") to all of its sentences by giving:

- A "domain of discourse" D (E.g.: a set)
- For every constant symbol, an element of D as its interpretation.
- For every n-ary function symbol α, an n-ary function $f_\alpha : D^n \to D$ as its interpretation.
- For every n-ary predicate symbol, a subset of D^n as its interpretation.

Mention Lowenheim-Skolem.

4. Peano Axioms

Succinctly, and distorted into its ready-at-hand (standard) model in set-theoretic language here are, in effect, Peano's axioms: Let $(N, 0)$ be a pointed set together with an injective mapping $S : N \hookrightarrow N$ such that $0 \notin S(N)$.

Say that $(N, 0)$ "is" a *form of the natural numbers* if for any subset $N' \subset N$ containing 0 and stable under the action of S, we have the equality $N' = N$. This latter requirement is

one way of formulating the induction axiom. This definition is "second-order"—e.g., there's a universal quantification over subsets: $\forall N' \subset N$— and its proof of "categoricity" is also "second-order." There are, as well, "first-order formulations" of "the" induction axiom. The quotation-marks around the definite article ("the" induction axiom) is to signal that it is really a list of axioms, one axiom (call it Ind_ϕ) 'personally designated' for each predicate ϕ taken from an appropriately set up language. Not to dwell on the fine points here, the axiom Ind_ϕ, in effect, is of the following form:

$$\text{Ind}_\phi : \quad (\phi(0) \wedge \forall x(\phi(x) \Rightarrow \phi(S(x))) \Rightarrow \forall x \phi(x)$$

But for this there is no categoricity: there are non-standard models of Peano axioms framed with the above "first-order form of induction," these models containing elements not correlated with any natural number[13].

Note that in the format of first-order induction $\phi \mapsto \text{Ind}_\phi$ described above, the language (in which these ϕ's are required to be expressed) allows for only a countable number of ϕ's, but there are uncountably many subsets of the natural numbers. So perhaps it isn't terribly surprising that there are nonstandard models of this axiom scheme.

The naive recursive strategy, piggybacking on a set theory though, achieves a measure of categoricity while giving up the ability (in a zero-th or first order logic) to talk about all natural numbers[14].

5. **Non-standard models** *E.g., from the compactness theorem:* The existence of non-standard models of arithmetic can be demonstrated by an application of the compactness theorem. To do this, a set of axioms P^* is defined in a language including the language of Peano arithmetic together with a new constant symbol x. The axioms consist of the axioms of Peano arithmetic P together with another infinite set of axioms: for each numeral n, the axiom $x > n$ is included. Any finite subset of these axioms is satisfied by a model which is the standard model of arithmetic plus the constant x interpreted as some number larger than any numeral mentioned in the finite subset of P^*. Thus by the compactness theorem there is a model satisfying all the axioms P^*. Since any model of P^* is a model of P (since a model of a set of axioms is obviously also a model of any subset of that set of axioms), we have that our extended model is also a model of the Peano axioms. The element of this model corresponding to x cannot be a standard number, because as indicated it is larger than any standard number.

6. **Gödel Numbering**.

[13]The operations of *Addition* and *Multiplication* and can be defined. E.g., Addition is the function $+ : N \times N \to N$ defined as: $0 + a = a = a + 0$ and $S(a) + b = S(a + b)$. Hence the Order relation "$\geq$" can also be defined: $a \geq b$ iff $\exists c \mid a + c = b$.

[14]The strategy is to work methodically, starting with the empty set, and by methodically invoking sets of subsets of sets built out of the two-some consisting of the empty set and the set containing the empty set as its only member, constructing an ordered system of finite sets that reifies the natural numbers. Of course, one will only be able to invoke—in any argument—only finitely nmany such natural numbers at any time—the totality being out of the first order language. However, this scheme is recursive and excludes the complications introduced when one needs to face the possibility of nonstandard models.

Discuss.

Part IV

Notes for the session: *Probabilistic models*

13 'Educating your beliefs' versus 'Testing your Hypotheses'

This is the handout about Bayesian issues for our session on Nov. 5, 2013.— B. Mazur

14 'Bayesian intertwining'

The naive view of an empirical investigation which we might call the **straight Baconian model** for a scientific investigation has, as recipe:

Set-up and Hypotheses $\longrightarrow$ **Data Collecting** $\longrightarrow$ **Processing Data and Conclusion.**

The manner in which one proceeds from data to conclusion is often understood to be a straight comparison of what the hypotheses would predict and what the data reveals[15], the comparison being (usually) quantitative with a pre-specified tolerance of discrepancy (between prediction and observation).

All this is significantly modified by the Bayesian viewpoint, which methodically intertwines the first two steps, and has a different take on each of these ingredients: hypothesis, data, conclusions. We'll discuss this below[16]. We'll look at the Bayesian viewpoint as offering a 'model' to help us understand, and deal with, the interplay between those ingredients. Let's call it the **Bayesian model** for a scientific investigation.

A further issue that complicates the contrast of *models of getting to scientific conclusions* alluded to above is the difference between the Bayesian's and the Frequentist's work; their methods are not

[15] although it might be difficult to find this expressed in Bacon's writings as bluntly

[16] A disclaimer: I know very little statistics; I'm a total outsider to this field and especially to the extended conversation—and the somewhat sharp disagreements—that Bayesians and Frequentists have.

the same, and they have slightly different *primary goals*.

We'll get to that, eventually.

15 Prior information and the Birthday problem

To introduce ourselves to this 'Bayesian intertwining' (taking as a **black box**—at least at first—some of the mathematical procedures involved) let's revisit a famous problem: the birthday problem. You have a class of fifth graders in an elementary school. Suppose there are 23 students in the class. What is the probability that two of them have the same birthday? Or, to seem more mathematical, suppose there are n students. What is the answer as a function of n?

Here is the simple naive analysis of this problem. We assume, of course, that the probability of anyone having a birthday at any specific day, e.g., April 22, is 1/365 (ignoring the leap year issue). Think of the teacher marking off—successively— on a calendar the birthdays of each student. We are going to gauge the possibility that in his class of n students there are no two birthdays on the same calendar day. The first student's birthday is duly marked. We can't possibly have a concurrence of birthdays (call it a *hit*) at this point, there being only one mark. So we can record "1" as the probability that we didn't get a *hit* at least so far[17].

As for the second student, the probability of him or her not having a birthday on the same day as student #1—i.e., that there not be a *hit*— is

$$1 - \frac{1}{365} \;=\; \frac{364}{365}.$$

Given this situation, and passing to the third student, in order for there not to be a hit, his or her birthday has to avoid two days, so that probability is

$$1 - \frac{2}{365} \;=\; \frac{363}{365}.$$

Putting the two probabilities together we get that–so far in our count—the probability that there isn't a hit with these three students is

[17] We are going to write probabilities as numbers between 0 and 1. So if the probability of an event is $\frac{1}{2}$ that's the same as saying that it is *even odds of it happening or not happening* or that 50% *of the time it happens*, or one sometimes simply says that there's is a 50/50 chance of it occurring.

$$(1 - \frac{1}{365})(1 - \frac{2}{365}) \quad = \quad (\frac{364}{365}) \cdot (\frac{363}{365}).$$

Working up (by mathematical induction) the probability that there's no hit, with n students is then:

$$(1 - \frac{1}{365})(1 - \frac{2}{365}) \cdots (1 - \frac{n-1}{365}),$$

which when $n = 23$ is close to $\frac{1}{2}$. That is, for a class of 23 students the chances are 50/50 that there's a concurrence of birthdays—given this analysis.

My Bayesian friend Susan Holmes tells me that she has actually tried this out a number of times in real live classes, and discovered that the odds seem to be much better than 50/50 for 23 students; you even seem to get 50/50 with classes of as low as 16 students.

There is something too naive in the analysis above, says Susan. We should, at least, make the following (initial) correction to our setting-up of the problem. We said above:

> We assume, of course, that the probability of anyone having a birthday at any specific day, e.g., April 22, is 1/365

BUT we actually *know* stuff about the structure of our problem that we haven't really registered in making that assumption.

For example, it is a class of fifth-graders so, chances are, they were all (or mostly) born in the same year. In particular, the years of their birth all (or mostly) had the same weekends and weekdays. In the era of possible c-sections and induced births—given that doctors and hospital staff would prefer to work on weekdays rather than weekends—one might imagine that the probability of being born on a weekday is somewhat skewed. We also know more that might make us think that fixing 1/365 at the rate is too naive.

Perhaps then, instead of sticking to the probability $p = 1/365$ per day hypothesis, allow a bit of freedom and a priori allow that there are different probabilities

$$p_1, p_2, p_3, \cdots, p_{365}$$

for each day of the year[18], about which we can make very very rough guesses. But let us not write this in stone yet. Make a mildly educated guess of these p_i; e.g., if "i" is a Saturday or Sunday (or a

[18] these summing to 1

holiday), then p_i is probably slightly less than 1/365; if a weekday, slightly more. This initial guess (of the values of $p_1, p_2, p_3, \cdots, p_{365}$) we'll call a **Prior**. From any prior we can deduce—essentially by a straight computation as we did above with the "constant prior: 1/365"—all the expected odds and whatever statistics one wants.

BUT we have hardly gotten our best answer! All these p_i's constituted, after all, just our very very rough guess based on some intuitive hunch, prior to having any hard data. Computing with these p_i's gives us a "number" as output—perhaps more accurate than the 23 we started this discussion with, but how does this number compare with the actual numbers we're actually accumulating by sampling birthday statistics for classes of fifth-graders? The Bayesian will use this accumulating *Data* to "correct" the prior (guessed) probabilities p_i, to be more in tune with the data. This is what I mean by the Bayesian intertwining: the data–as it comes in–is used to "educate the prior." This educated-prior is called (naturally) a **posterior**. In some sense, the principal role of data in this Baysian model is *to be fed back into the prior to refine it* to produce successive posteriors rather than (with a straight up or down judgment) to verify or contradict an hypothesis.

Starting anew with the latest **posterior** rather than the original **prior** we can deduce—essentially as we did above with the "constant 1/365" or any prior **prior** —all the expected odds and whatever statistics one wants.

In fact, there are no firm hypotheses within the Bayesian model, and no firm conclusions. I said, though: "*within* the Bayesian model." Nevertheless from this procedure one might extract a conclusion, but this is outside the format.

This is a preliminary move in the Bayesian direction, but we aren't quite there yet. Another–and better–way of viewing this move (reflecting our most up-to-date version of belief about the set-up) is that the initial values

$$p_1, p_2, p_3, \cdots, p_{365}$$

should not be taken as hard unchangeable numbers but rather are to be viewed as "random variables" in their own right, and subject to their own distributions, which we are bent on determining, given enough **Data**. The grand function of the data is to be fed back to educate the prior but retaining its status as probabilities. The movement here is as follows:

$$\text{Prior (probabilities)} \quad \overset{\textbf{Data}}{\longrightarrow} \quad \text{Posterior (probabilities).}$$

The **black box**—so far—is that I have not yet said anything about the mathematical procedure Bayesians use to feed back (as an afterburner) information obtained by the Data into the prior

assumptions, in order to effect the "education" of these prior assumptions and thereby produce the posterior. For the moment—in this discussion—it is more important for me simply to emphasize that *whatever this procedure is* it is, in fact, a *predetermined procedure.*

16 Predesignation versus the self-corrective nature of inductive reasoning

Now you might well worry that this Bayesian ploy is like curve-fitting various hypotheses[19] to the data—a kind of hypothesis-fishing expedition, if you want. You keep changing the entire format of the problem, based on accumulating data. The Bayesians have, as I understand it, a claim: that any two 'reasonable' priors, when "corrected" by enough data will give very close posteriors. That is, the initial rough-hewn nature of the prior will iron out with enough data. Their motto:

> *Enough data swamps the prior.*

I've been playing around with another formulation of that motto:

> *Any data-set is, in fact, a 'data point' giving us information about the probability distribution of priors.*

In contrast, there is a motto that captures the sentiment of a Frequentist:

> *Fix hypotheses. This determines a probability distribution to be expected in the data. Compute data. If your hypotheses are good,* **in the limit** *the data should conform to that probability distribution.*

About the above, one of the early great theorizers in this subject (and specifically regarding probability, randomness, and the law of large numbers) was Jacob Bernoulli. He *also* was a theologian preaching a specifically Swiss version of Calvinism. You see the problem here! There is a strict vein of *predetermined* destiny or fatalism in his theology, someone who is the father of the theory of randomness. How does he reconcile these two opposites? Elegantly, is the answer! He concludes[20] his treatise *Ars Conjectandi*, commenting on his law of large numbers, this way:

[19] I want to use the word *hypothesis* loosely, for the moment; that is, the way we generally use the word; and not in the specific manner that statisticians use it.

[20] It is, in fact, the conclusion of the *posthumously* published treatise (1713) but it isn't clear to me whether or not he had meant to keep working on the manuscript.

> Whence at last this remarkable result is seen to follow, that if the observations of all events were continued for the whole of eternity (with the probability finally transformed into perfect certainty) then everything in the world would be observed to happen in fixed ratios and with a constant law of alternation. Thus in even the most accidental and fortuitous we would be bound to acknowledge a certain quasi necessity and, so to speak, fatality. I do not know whether or not Plato already wished to assert this result in his dogma of the universal return of things to their former positions [apokatastasis], in which he predicted that after the unrolling of innumerable centuries everything would return to its original state.

Apokatastasis is a theological term, referring to a return to a state before the fall (of Adam and Eve)[21].

17 Priors as 'Meta-probabilities'

Suppose you are a cancer specialist studying a specific kind of cancer and want to know if there is a gender difference: do more men than women get this type of cancer? Or more women than men?

Now suppose I asked you (cancer specialist) to make some kind of guess—when considering groups of people that get this cancer—about the proportion of men-to-women that get it. You might tabulate this as a probability P that a random choice of person in this group is male. So P is a number between 0 and 1. You might actually give me a number if you are very confident, but more likely, for a spread of possible values of P, you'll give me an estimate of greater or lesser levels of confidence you have that this P is indeed the sought-for-probability. Taking the question I asked more systematically, you might interpret it as follows:

> As P ranges through all of its possible values, from 0 (no males get it) to 1 (only males get it) tell me (your guess of) the probability that P is the ratio $\frac{M}{M+W}$ where M is the

[21] Noah Feldman once suggested to me that Calvinists might be perfectly at home with random processes leading to firm limiting fatalism, in that the fates of souls—in Calvinist dogma—are *randomly assigned* and not according to any of their virtues; i.e., to misquote someone else: "goodness had nothing to do with it."

Also, we might connect the above with C.S. Peirce's 1883 paper "A Theory of Probable Inference." For a readable discussion of this paper, see: Len O'Neill's *Peirce and the Nature of Evidence* published in the Transaction of the Charles S. Peirce Society **29** Indiana Univ. Press (1993) pp. 211-224. Peirce makes a distinction between *statistical deduction* and *statistical induction* the first being thought of as reasoning from an entire population to a sample, and the second being reasoning from sample to population. As O'Neill says, in the first it is a matter of long run frequency (i.e, the Frequentist's motto) whereas the second is related to a Peircean conception of *the self-corrective nature of inductive reasoning* (and this sounds like the Bayesian protocol).

Peirce dwells on the issue of *predesignation* in the Frequentist's context (i.e., you fix a model and then collect evidence for or against it; you don't start changing the model midstream in view of the incoming evidence). As already mentioned, there is a curious type of *meta-predesignation* in the Bayesian context, in that the manner in which you change the model, given incoming evidence, is indeed pre-designated.

number of men and W the number of women in the group? In effect, draw me a graph telling your probability-estimate for each of the P's in the range between 0 and 1.

Your initial guess, and initial graph, is the Prior (I privately call it the *meta-probability*). It *will* be educated by the data accumulating.

Let's imagine that you say "I have no idea! This probability P could–as far as I know–equally likely be any number between 0 and 1." If so, and if you had to draw a graph illustrating this noncommittal view, you'd draw the graph of a horizontal line over the interval $[0, 1]$. Or, you might have some reason to believe that P is close to $1/2$ but no really firm reason to believe this and you might have no idea whether gender differences enter at all. Then the graph describing your sense of the likelihood of the values of P would be humped symmetrically about $P = 1/2$. Or if you are essentially certain that it is $1/2$ you might draw it to be symmetrically spiked at $P = 1/2$.

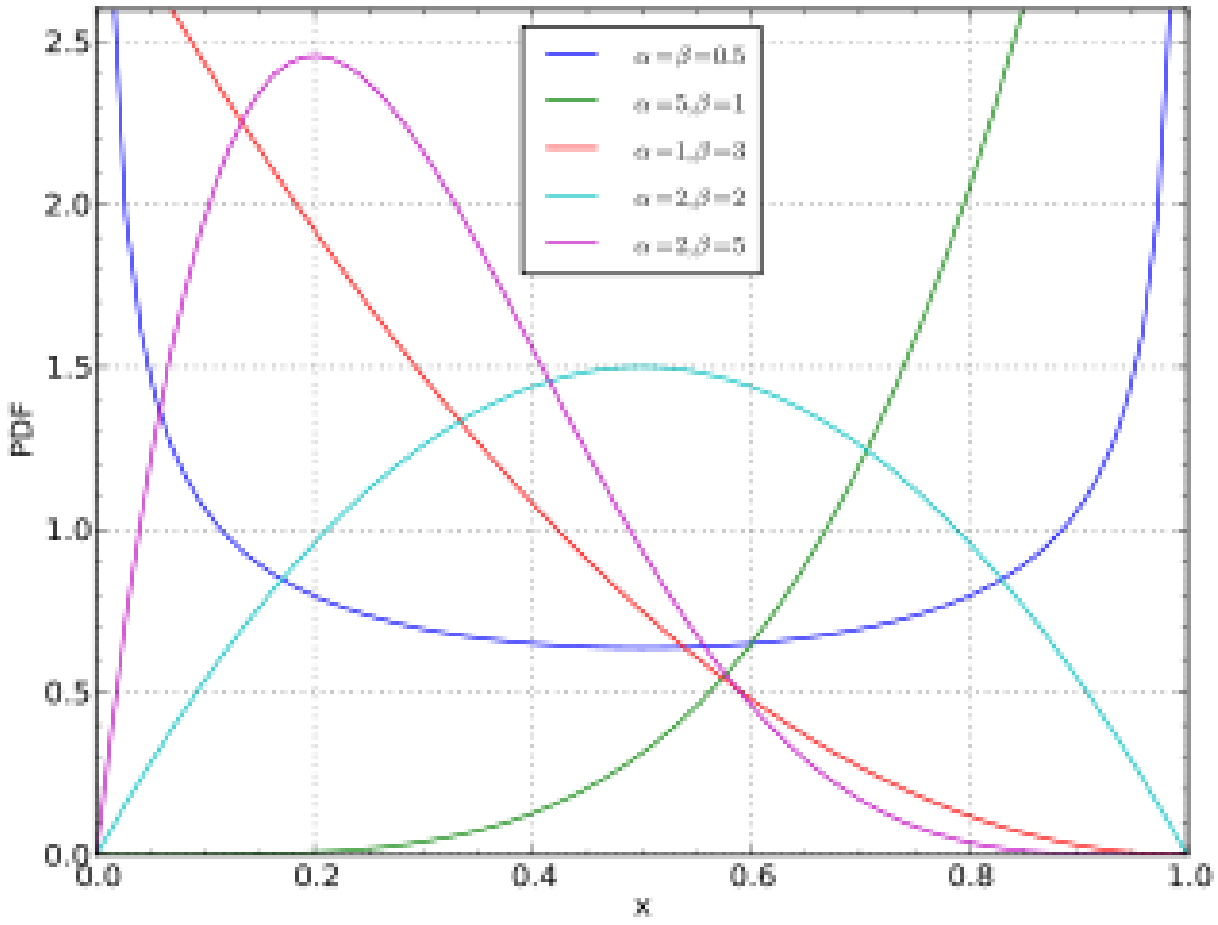
What you are drawing is–in a sense–a *meta-probability density* since you are giving a portrait of your sense of how probable you think each value between 0 and 1 might be the actual probability-that men-get-this-type-of-cancer. Your portrait is the graph of some probability density function $f(t)$.

There are theoretical reasons to suggest, for some such problems, that you would do well to be drawing the graphs of a specific well-known family called **beta-distributions**. These beta-distributions come as a two parameter family[22] $\beta_{a,b}(t)$. That is, fix any two positive numbers a, b (these numbers a, b are called the *shape parameters* of the beta-distribution) and you get such a graph.

Here are some general ground-rules for choosing these βs: shape parameters that are equal give distributions symmetric about $1/2$; i.e., you choose such a β if you expect that gender plays no role in the probability of contracting this cancer. Choosing $a > b$ means that you are skewing things to the left; i.e., you believe that men get this type of cancer less frequently than women; choosing $b > a$ means the reverse. The larger these parameters, the sharper the peak of the curve; i.e., the more "sure" you are that the probability occurs at the peak.

Choose parameters, say, $a = 2, b = 5$; or, say, $a = 2, b = 2$ and you have probability distributions $\beta_{2,5}(t)$, or $\beta_{2,2}(t)$, these being the blue and the magenta graphs in the figure below (taken from a wonderful Wikipedia entry: `http://en.wikipedia.org/wiki/Beta_distribution`).

[22] These are distributions $t^{a-1}(1-t)^{b-1}dt$ normalized to have integral equal to 1 over the unit interval.



18 Back to our three steps

1. (**Choosing the Prior**) Now, Bayesian cancer doctor that you are, when you start doing your statistics, choose a Prior. For this type of question you might do well, as I said, to choose some beta-distribution. If you imagine that there might be a gender bias here, but have no idea in which direction, you might choose one that is symmetric about $t = 1/2$ (which, as it turns out, means that you'd be taking shape parameters a equal to b). But size up the situation as best as you can, taking into account everything that you think is important to the problem and come up with a choice of a Prior. Let us say that your Prior is $\beta_{a,b}(t)$.

2. (**The Data**) Suppose you now get a data sample of 100 people with cancer—perhaps the result of some specific study of some particular population, and suppose that 60 of these cancer victims are men (so 40 are women).

3. (**Passing to the Posterior**) The beauty of the family of beta-distributions is that when you appropriately *educate* a beta-distribution (the Prior) with new data, the new distribution (the Posterior) is again a beta-distribution. The only thing is that the shape parameters may change; say, from (a, b) to a new pair of numbers (a', b'):

$$\beta_{a,b}(t) \quad \stackrel{\text{new data}}{\longrightarrow} \quad \beta_{a',b'}(t)$$

I'm told that this change can be very easily computed. That is, in this example problem, the a', b' will depend on hardly more than the original a, b, the percentage of men with cancer, and the size of the study.

19 A numerical example and a question

For this example I'm normalizing things so the numbers work simply so we don't get bogged down in mere arithmetic. Imagine that your Prior is $\beta_{20,20}$ and you test a sample population (of just the right size for the normalizations to work out as I'm going to assume they do below) and in that population Men/ Women cancer ratio is 60/40. The Posterior is then (I'm told) $\beta_{20+60,20+40}$. And if you compute (based on that Posterior) the probability that men get this type of cancer more than women, that probability is:

$$0.955\ldots$$

If you did the analogous thing with the Prior $\beta_{10,10}$, getting, as Posterior, $\beta_{10+60,10+40}$ you'd compute (based on that Posterior) the probability that men get this type of cancer more than women to be:

$$0.966\ldots$$

Question: Why is it *reasonable* that the second estimate of probability of gender-difference be bigger than the first?

20 Bayes' Theorem

I will assume that people have learned something of the background of Bayes himself from the other readings, and just concentrate on the statement and intent of his theorem. (A discussion of his theorem often is the *start* of expositions on Bayesian things, but it seems to me that one needs some of our prior discussion if we want (not merely to understand the theorem, but) to focus on its effect: a distinctive model of the role of statistical inference in the formation of scientific conclusions.

To begin, imagine that we have a region C of the Euclidean plane (say, an open compact subset) and two (open) subsets of C, $X, Y \subset C$, so that we also may consider their intersection $X \cap Y$ viewed as a subset of X and of Y.

Put figure here

Letting the absolute value sign $|\ |$ indicate area, we have the tautology:

$$(*) \qquad \frac{|X \cap Y|}{|Y|} \cdot \frac{|Y|}{|C|} \;=\; \frac{|X \cap Y|}{|X|} \cdot \frac{|X|}{|C|}.$$

This is evident.

Now, interpret C as a community of individuals or entitities, and X, Y as the sub-communities of C consisting of individuals that have specific traits (call them, respective, 'trait x' and 'trait y'). View 'area' as giving numbers of individuals.

Then:

- we can think of $\frac{|X \cap Y|}{|Y|}$ as the *probability that an individual in the community C has trait x, given that it has trait y.* This is usually abbreviated: $P(x \mid y)$.
- we can think of $\frac{|X \cap Y|}{|X|}$ as the *probability that an individual in the community C has trait y, given that it has trait x.* This is usually abbreviated: $P(y \mid x)$.
- we can think of $\frac{|X|}{|C|}$ as the *probability that an individual in the community C has trait x, abbreviated as $P(x)$*, and
- we can think of $\frac{|Y|}{|C|}$ as the *probability that an individual in the community C has trait y, abbreviated as $P(y)$.*

Rewriting the tautology **(*)** above in terms of these probabilities "P" we have:

$$(**) \quad P(x \mid y) \cdot P(y) \;=\; P(y \mid x) \cdot P(x).$$

interpreted as:

> The conditional probability that an individual in the community C has trait x, given that it has trait y *times* the probability that an individual of the community has trait y
>
> is equal to
>
> the conditional probability that an individual in the community C has trait y, given that it has trait x *times* the probability that an individual of the community has trait x.

This is Bayes' Theorem, which is sometimes written:

$$(***) \quad P(x \mid y) \;=\; \frac{P(y \mid x) \cdot P(x)}{P(y)}.$$

So, what does this theorem have to do with the discussion we've given in the previous sections of this handout? The answer (in my opinion) has two prongs.

- I think that the more important connection that Bayes' Theorem has to the general Bayesian viewpoint is that the theorem is a 'promissory note,' so to speak, that *conditional probabilities*—i.e., probabilities based on conditions that express what things we know about the situation—will be our vocabulary, and we have the beginnings of a way of dealing with conditionality.

- But it also has a 'straightforward' type of application. Here's one of enumerably many such simple examples (all below gotten from Wikipedia).

  Suppose you are an entymologist dealing with a species of beetles and there is a rare subspecies of beetle, usually identifiable because of a certain pattern on its back. I say, usually, but not always. Here is what you know:

  – the probability[23], given that you have an individual of the rare subspecies, that the pattern occurs on its back is 98%. I.e.,

  $$P(Pattern|Rare) = 98\%,$$

  and

[23] I'll give probabilities here in terms of percentages.

– the probability, given that you have a 'common' individual of the species, that the pattern occurs on its back is 5%. I.e.,

$$P(Pattern|Common) = 5\%.$$

Moreover, the rare subspecies accounts for only 0.1% of the entire population of this species of beetle.

Now you capture a beetle with the pattern on its back. What is the probability that it is a member of the rare subspecies? That is, what is $P(Rare|Pattern)$? Bayes Theorem comes to the rescue.

For further concrete applications, see `http://en.wikipedia.org/wiki/Bayes'_theorem` which is the first (and probably the most useful) of the many hits you get when you Google "Bayes' Theorem."

Part V

Notes to the session: *Model Theory in Mathematics, and Models per se*

21 Introduction

In contrast to the other three sessions devoted to specific connections between Mathematics and Models—the first, an introductory survey of applied mathematical models, the second of axiomatic models, and the third devoted to the Bayesian statistical viewpoint—in this fourth session I would like to have two different discussions (given enough time).

The main part of our session will have to do with the notion of "model" internal to mathematics—i.e., Model Theory. I would also like a—however brief— second discussion dealing with the strange (novel) notion that the ease of accumulating 'Big Data' puts us beyond the task of actually formulating precise models.

22 The Lowenheim-Skolem Theorem

Thoralf Skolem wrote (at least) two articles about this theorem, both reprinted in **[F-G]** *From Frege to Godel: A Source Book in Mathematical Logic,* 1879-1931 Edited by Jean van Heijenoort:

- **AST:** *Axiomatized Set Theory* pp.290-301 of [**F-G**],
- **LCI:** *Logico-Combinatorial Investigations* pp.252-263 of [**F-G**].

Even before discussing the terms here, it could be illuminating to read the *Concluding Remarks* of [**AST**] (pp. 300 bottom to 301).

> The most important result above is that set-theoretic notions are relative...I believed that it was so clear that axiomatization in terms of sets was not a satisfactory ultimate foundation of mathematics that mathematicians would—for the most part—not be very much concerned with it. But in recent times I have seen to my surprise that so many mathematicians think that these axioms of set theory provide the ideal foundation for mathematics; therefore it seemed to me that the time had come to publish a critique.

Almost every concept of modern mathematics depends on the vocabulary of sets. There is something rock-bottom and primitively pre-mathematical when we think about the combinatorics of finite sets. So one would naively think that a genuine theory of sets (infinite ones included), in complete generality, has a firm definiteness. And ubiquity.

After all, the general 'model' of any mathematical theory has the neat format and the bare structure of the theory of *categories*: a **Category** is a structure that consists of *objects* and *morphisms* (i.e., mappings from X, one object of the category, to Y, another). The 'collection' of *objects* of the category is given the status of a 'class' (whatever that is...) and–at least in most day-to-day applications—the morphisms from an X to a Y is given the structure of a set.

But here, in Skolem's article [**AST**] (and in the other articles in [**F-G**]) one sees a tug of war between

- the strict formalization of Set Theory with a capital "S"

and

- the realization that the formalization (e.g., the collection of Zermelo's axioms) itself allows one to have a somewhat relative attitude toward the nature of what "a" set theory is.

That is, the axioms simply work with a domain of objects and relationships, where it is the interrelationships between the various aspects of this structure that is axiomatized, and therefore the structure can be viewed as possibly having many 'realizations,' or models. This is very much like one we have axiomatized *group theory*, say, we can search for quite different realizations of those axioms—i.e., different groups.

It is interesting, with this in mind, to read footnote 9 (p.209) of [**AST**], written in 1922—which explicitly guesses the result of the 1938 theorem of Gödel and the 1964 theorem of Cohen. These theorems taken together prove the independence of the continuum hypothesis. The modus operandi of Cohen is *forcing* which seems already hinted at in this footnote.

The Lowenheim-Skolem Theorem expresses this type of relativity in more fundamental terms.

Every first-order proposition is either a contradiction or else it is satisfiable in a finite, or denumerably infinite domain.

Here is Skolem's description of how this can be used to interpret Zermelo's axioms:

> The definite first-order propositions can be enumerated according to their form by means of the positive integers, so that these in turn can be thought of as ordered according to some rule. Consequently, Axiom III (axiom of separation) can be replaced by an infinite sequence of simpler —which, like the rest of Zermelo's axioms, are first-order propositions in the sense of Löwenheim—containing the two binary relations $\in$ and $=$. We may conclude: *If Zermelo's axiom system, when made precise, is consistent, it must be possible to introduce an infinite sequence of symbols* $1, 2, 3, \ldots$ *in such a way that they form a domain* B *in which all of Zermelo's axioms hold provided these symbols are suitably grouped into pairs of the form* $a \in b$. This is to be understood in the following way: one of the symbols $1, 2, 3, \ldots$ will be the null set (that is, among the remaining symbols there is none that has the relation $\in$ to the symbol in question); if a is a symbol, then $\{a\}$ is another; if M is one of the symbols, then UM, SM and DM are others; and so forth.
>
> So far as I know, no one has called attention to this peculiar and apparently paradoxical state of affairs. By virtue of the axioms we can prove the existence of higher cardinalities, of higher number classes, and so forth. How can it be...?

Compare: `http://en.wikipedia.org/wiki/Skolem's_paradox`

The technique that Skolem uses to get his discussion going (at least in his slightly easier paper [**LCI**]) is three-step:

- **Replace first-order propositions by equivalent ones in 'normal form.'**

  Skolem notes that the general mess of a first-order proposition that has a syntax within a certain language and takes its terms from within a domain D of entities has the form

  $$\forall_{x_1}\exists_{y_1}\forall_{x_2}\exists_{y_2}\forall_{x_3}\exists_{y_3}\ldots\forall_{x_n}\exists_{y_n}P(x_1, x_2x_3, \ldots, x_n; y_1, y_2, y_3, \ldots, y_n)$$

  and that he can replace any such conglomerate by the simpler *normal form*

  $$\forall_{x_1}\forall_{x_2}\forall_{x_3}, \ldots, \forall_{x_n}\exists_{y_1}\exists_{y_2}\exists_{y_3}, \ldots, \exists_{y_n}P(x_1, x_2x_3, \ldots, x_n; y_1, y_2, y_3, \ldots, y_n)$$

  or equivalently,

  $$\forall_{x_1, x_2x_3, \ldots, x_n}\exists_{y_1, y_2, y_3, \ldots, y_n}P(x_1, x_2x_3, \ldots, x_n; y_1, y_2, y_3, \ldots, y_n)$$

or even more succinctly, allowing $\mathbf{x}$ and $\mathbf{y}$ to stand for the evident n-tuples, we can write it as

$$\forall_{\mathbf{x}} \exists_{\mathbf{y}} P(\mathbf{x}, \mathbf{y}).$$

- **Use of Axiom of Choice.**

  Recall that these $\mathbf{x}$'s and $\mathbf{y}$'s are "tuples" of elements from the domain D. Now, given any first-order proposition P as above, and granted its truth, for any $\mathbf{x}$, choose a $\mathbf{y} := \mathbf{x}'$ such that $P(\mathbf{x}, \mathbf{x}')$. We can then replace the gadgets $\forall_{\mathbf{x}} \exists_{\mathbf{y}} P(\mathbf{x}, \mathbf{y})$ by

  $$\forall_{\mathbf{x}} P(\mathbf{x}, \mathbf{x}').$$

- **Start appropriately and keep going.**

  Enumerate all the first-order propositions in normal form in your language and starting with the basic (finite, or denumerable) subdomain $D_0 \subset D$ that you need just to get started, and keep augmenting it to $D_0, D_1, D_2, \ldots$ where at the n-th stage the elements in the $\forall$'s are—inductively—taken from D_n allowing you to build a D_{n+1} so as to systematically include the entries of the tuples $\mathbf{x}'$ coming from those $\mathbf{x}$'s. When you are done, you get a denumerable (or finite) subdomain $D_\infty \subset D$ within which, when you interpret your first-order propositions, exactly the ones true for D are the ones true for D_∞ (and ditto for the false ones).

23 Making models within models.

Although the above is the format for engineering denumerable domains that mimic the first-order behavior of true sentences in possibly 'larger' domains, it remains to puzzle out the seeming paradox that Skolem underlines: "By virtue of the axioms we can prove the existence of higher cardinalities, of higher number classes, and so forth. How can it be?"

One response to this (as Skolem hints obliquely) is that the underlying notion of *cardinality* is neatly amenable to violent relativization. Try this: Imagine that you have a perfectly nice (call it a 'starter') set theory which allows for only finite or denumerably infinite sets. So, in fact, in your starter set theory, any two (denumerably) infinite sets are in one:one correspondence.

Now, the notion of cardinality depends on one:one correspondence. A one:one correspondence between two sets X and Y is given by some graph—i.e., a subset of the product.

Imagine then that you also have $\mathcal{C}$: a collection of axioms and propositions that give a first-order account of some set theory allowing for finite sets, denumerable sets, and at least some higher cardinals; so it may be including something we can call $\aleph_1$, whatever that is. Although the procedure for giving a denumerable model for the first-order theory of this $\mathcal{C}$ is bewilderingly artificial, imagine that procedure to have 'modeled' $\mathcal{C}$ in starter set theory in the following *mise en abyme* manner: the model of $\mathcal{C}$ is given by stipulating an artful collection of sets of the starter theory $X, Y, \ldots$.

Now, even though X and Y may *have been* both infinitely denumerable (and therefore in one:one correspondence with each other) in the starter theory, imagine that the model of $\mathcal{C}$ contains none of the subsets of their product $X \times Y$ corresponding to graphs of one:one correspondence between X and Y. I.e., *in the perspective of this model of* $\mathcal{C}$, X and Y are no longer of the same cardinality. But imagine, as well, that the subset of $X \times Y$ that corresponds to an embedding of X in Y is part of the model. Thus, in $\mathcal{C}$ modeled in the starter set theory, Y would have strictly larger cardinality than X, even though in the starter theory these are both denumerable.

Although I haven't looked into the history of the contemporary subject, *Model Theory*, it seems likely to me that these papers of Skolem (and of Lowenheim before him) constituted the launching of the model-theoretic viewpoint.

But it most certainly launched a fascinating epistemological discussion proposing a certain meta-relativity having to do with 'reality,' where (as in the 'Brains in a vat' chapter of Hilary Putnam's *Reason, Truth and History* written in the late 70s) the Lowenheim-Skolem theorem is offered as a metaphorical format for 'Truth.' Putnam comments on the then current philosophical discussions that seemed to be forced into the two extreme poles of a dichotomy: an 'objective' view (sometimes referred to as a 'copy theory,' where—to put it bluntly–there are facts and truth copies those facts) and a 'subjective' view (in Putnam's words: a "hopelessly subjective view," taken by Kuhn in his most outspoken moments, and by Feyerabend, and Foucault). Putnam then offers his synthesis of the objective and subjective—mirroring a Skolem-Lowenheim picture.

24 Model Theory

The task for our discussion here is simply to get a sense of what a model–in the sense of modern Model Theory— means and what it is intended to clarify. To have our discussion somewhat specific, please make sure that you have—in your readings—come across these terms, as they are construed in Model Theory. For this,

1. Read pp. 7-21 in Stephen Simpson's *Model Theory*. This is a gentle introduction to the subject. The full text can be gotten on line `http://www.personal.psu.edu/t20/notes/master.pdf`.

2. Compare the above with pp. 15-28 in `http://library.msri.org/books/Book39/files/marker.pdf` (which is the first chapter in `http://library.msri.org/books/Book39/files/fm1.pdf`. Thanks to Rehana Patel for mentioning the on-line errata for Marker's text: `http://homepages.math.uic.edu/~marker/mt-errors.html`.

We follow Simpson in the glossary below:

- **Symbols:**

- nonlogical symbols: n-ary relation symbols $R(*,*,\ldots,*)$; n-ary operation symbols $o(*,*,\ldots,*)$; constant symbols c.
- logical symbols: for propositional connectives (negation, conjunction, disjunction, implication, bi-implication); quantifiers (the universal $\forall$, the existential $\exists$); equality"="; variables.

- **Term:** This is defined inductively as follows. A constant symbol is a term. A variable is a term. If $t_1,\ldots,t_n$ are terms and o is an n-ary operation symbol, then $o(t_1,\ldots,t_n)$ is a term.

- **Atomic formula:** This is defined as follows. If t_1 and t_2 are terms, then $t_1 = t_2$ is an atomic formula. If If $t_1,\ldots,t_n$ are terms and R is an n-ary relation symbol, then $R(t_1,\ldots,t_n)$ is an atomic formula.

- **Formula:** This is defined inductively as follows. An atomic formula is a formula. If ϕ and ψ are formulas then so are
$$\neg\phi, \phi\wedge\psi, \phi\vee\psi, \phi\to\psi, \phi\leftrightarrow\psi.$$
If ϕ is a formula and v is a variable, then $\forall v\phi$ and$\exists v\phi$ are formulas.

- **Sentence:** A **sentence** is a formula with no free variables.

- **Structure:** A **structure** is an ordered pair $A = (|A|, \phi)$ where $|A|$ is a nonempty set, called the **universe** of A, and ϕ is a function whose domain is a set of non- logical symbols.

  The domain of ϕ is called the signature of A[24].

  Here is a hint about the range of ϕ: To each n-ary relation symbol $R \in sig(A)$ we assume that ϕ assigns an n-ary relation $\phi(R) \subset |A|^n = |A|\times\ldots\times|A|$. To each n-ary operation symbol $o \in sig(A)$ we assume that ϕ assigns an n-ary operation $\phi(o) : |A|^n?|A|$. To each constant symbol $c \in sig(A)$ we assume that ϕ assigns an individual constant $\phi(c) \in |A|$.

  For example: the *structure* of the real numbers can be given by taking $|A| := \mathbf{R}$, its signature is $\{+,?,,0,1,<\}$, and the "ϕ" sends each of these nonlogical symbols $+,?,,0,1,<$ to what you guess it sends them to.

- **Truth** In notation, at least, if somewhat oblique, we write:
$$A \models \sigma \quad (A \text{ satisfies } \sigma; \sigma \text{ is } \textit{true} \text{ in } A).$$
As Simpson writes: "For example,
$$R \models \forall x\exists y(y\cdot y = x \vee y\cdot y = -x)$$
expresses the fact that every real number or its negative is a square."

[24]If S is *any* set of formulas and/or terms, the **signature** of S is the set of all nonlogical symbols occurring in it.

- **Models:** Let S be a set of sentences. A **model** of S is a structure M such that $M \models \sigma \forall \sigma \in S$, and $sig(M) = sig(S)$.

  For example, a group can be described as a model

  $$\mathcal{G} = (|\mathcal{G}|, , -1, 1)$$

  of the axioms of group theory:

  $$\forall x \forall y \forall z((x \cdot y) \cdot z = x \cdot (y \cdot z)) \forall x(x \cdot 1 = 1 \cdot x = x) \forall x(x \cdot x^{-1} = x^{-1} \cdot x = 1).$$

The above discussion formulates (following Simpson's text) something that I want to call *Naive Model Theory*, where the adjective "naive" is meant in an approving rather than pejorative sense (e.g., compare it to Paul Halmos's classic expository text "*Naive Set Theory*). The naiveté of it is that its notion of truth depends on some ambient semantics that is undiscussed. This is, of course, in the end always necessary, but there are other (slight variant) takes on it. For example, Marker's text which is on the reading list, and whose vocabulary list I trust people have looked at and thought about. There, one is modelling (whatever is being modeled) *in a set theory*.

25 Big Data versus models

Regarding the strain on models posed by experimental mathematics and massive computation, Peter Norvig, Google's research director, pronounced the following dictum:

> "All models are wrong, and increasingly you can succeed without them."

This sentiment was taken up by Chris Anderson in a recent issue of *Wired* Magazine and extensively commented on by the mathematician George Andrews[25].

Anderson noted that traditional science depended on model-formation. He went on to say:

> The models are then tested, and experiments confirm or falsify theoretical models of how the world works. This is the way science has worked for hundreds of years. Scientists are trained to recognize that correlation is not causation, that no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you must understand the underlying mechanisms that connect the two. Once you have a model, you can connect the data sets with confidence. Data

[25] G. Andrews *Drowning in the Data Deluge*, Notices of the American Mathematical Society, **59** August 2012 (933-941).

without a model is just noise. But faced with massive data, this approach to science hypothesize, model, test is becoming obsolete. Consider physics: Newtonian models were crude approximations of the truth (wrong at the atomic level, but still useful). A hundred years ago, statistically based quantum mechanics offered a better picture but quantum mechanics is yet another model, and as such it, too, is flawed, no doubt a caricature of a more complex underlying reality. The reason physics has drifted into theoretical speculation about n-dimensional grand unified models over the past few decades (the "beautiful story" phase of a discipline starved of data) is that we don't know how to run the experiments that would falsify the hypotheses: the energies are too high, the accelerators too expensive, and so on. Now biology is heading in the same direction... In short, the more we learn about biology, the further we find ourselves from a model that can explain it. There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot [26].

Read: P. Norvig, *Colorless green ideas learn furiously; Chomsky and the two cultures of statistical learning* pp. 30-33 in *Significance* The Royal Statistical Society, August 2012. (This is in the attachment:"Norvig.Significance.pdf")

See also:

- G.E. Andrews, *The Death of Proof? Semi-Rigorous Mathematics? Youve got to be kidding!* Mathematical Intelligencer **16** (1994) 16-18.
- J. Borwein, J., P. Borwein, R. Girgensohn, S. Parnes, *Making Sense of Experimental Mathematics*, Mathematical Intelligencer **18** (1996) 12-18.

and the two books:

- J. Borwein and D. Bailey, *Mathematics by experiment: plausible reasoning in the 21st century*, AK Peters (2003)
- J. Borwein, D. Bailey, R. Girgensohn, *Experimentation in mathematics: computational paths to discovery*, AK Peters (2004)

For discussions of similar issues, but with attention paid to changes in attitudes towards foundations, see

[26] I want to thank Stephanie Dick who provided me with the following three references regarding a related topic; namely *automatized proofs* in mathematics.

- Donald MacKenzie, *Computing and the culture of proving*, Philosophical Transactions of the Royal Society **363** (2005) 2335-2350.
- Donald MacKenzie, *Slaying the Kraken: The sociohistory of mathematical proof*, in *Social Studies of Science* **29** (199) 7-60.
- Hao Wang, *Toward mechanical Mathematics*, IBM Journal (January, 1960) 2-22.

- Leo Corry's *The Origins of Eternal Truth in Modern Mathematics: Hilbert to Bourbaki and Beyond*, in *Science in Context*, **12** (1998) 137-183.

- D. A. Edwards and S. Wilcox's *Unity, Disunity and Pluralism in Science* (1980) arxiv.org/pdf/1110.6545

- Arthur Jaffe's *Proof and the Evolution of Mathematics*, in *Synthèse* **111** (2) (1997) 133-146.

.