

Statistical and Game Theoretic Representations

of States of the World and Information

Lecture 2: Introduction

(U. Bayreuth, Spring 2006)

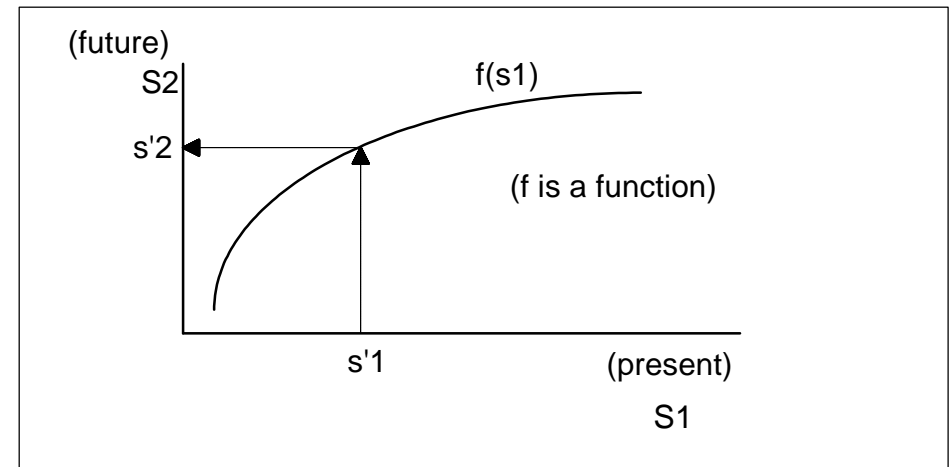
Last time we talked about the "informational" nature of experience, the idea of received wisdom about categories, language and norms. I argued that all experience is really information and all decisions are informational. (Choice comes before action.) And, I argued that it is possible to know something about the world. Insofar as communication is possible, it implies that we share a common information set. However, everyone's private collection of information is a bit different because experience and training tends to be just a bit different. Thus, there is also a sense in which we each live in our "own" unique world, even though that world has a lot in common with others.

A good deal of that variation occurs because of chance or random events. Today I want to discuss mathematical and statistical methods for representing the world in general and information in particular.

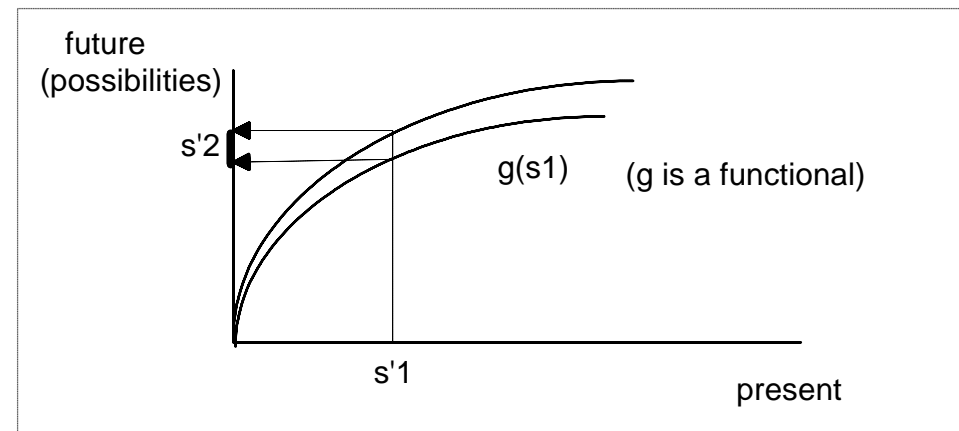
I. To think about the world using mathematical terms, we might first imagine a state of the world, S_1

- A. Suppose that we are completely confident that S_1 is the true state of the world, and that we are interested in what happens next, which we can call S_2 .
- B. If there is a **completely deterministic causal connection** between S_1 and S_2 , we can represent this with a function. $S_2 = f(S_1)$.
 - ♦ A function maps every point in its domain to a unique, single point, in its ranges.

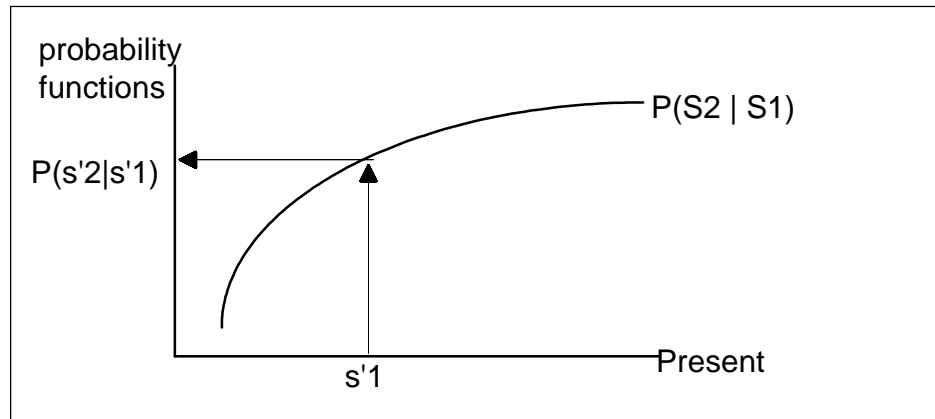
- ♦ In this case the world is like a great clock and one event (one cog in the clocks gear) necessarily leads to the next event (next cog in the gear).



- ♦ In such a world, there is no free will and there need be no uncertainty.
- C. If, however, there is stochastic or **random causal connection** between S_1 and S_2 , we can represent this with a *functional*. $S_2 = g(S_1)$.
- ♦ A functional maps particular points in a domain into range that consists of **sets** rather than single points. Here S_2 is a set rather than a single point.



- ♦ (Purists will note that a function is a special case of a functional, why?)



- D. A **conditional probability function** is a functional, that is slightly different from this; rather than mapping S_1 into what can happen, S_2 , it maps into a probability function over those possibilities, P_2 .
- ♦ $P_2 = f(S_2 | S_1)$ is the probability function over S_2 possibilities in period two.
 - ♦ That is to say, P_2 is the "conditional probability function" that characterizes the future given S_1 has happened and S_2 is the set of what could possibly happen given that S_1 has happened.
 - ♦ (In such a world, there could be free will. If your thought process has a random component, you will not be completely predictable even to yourself.)
- E. Notice that a conditional probability function may represent a real phenomena.
- ♦ For example if S_1 is "you roll a dice that has the numbers 1 to six on it", P_2 is $1/6$ for 1, $1/6$ for 2, $1/6$ for 3 and $1/6$ for 6 if the dice is fair.
 - ♦ If instead, if S_1 is "you roll a dice that has the even numbers 2 to 12 on it", P_2 is $1/6$ for 2, $1/6$ for 4, $1/6$ for 6 and $1/6$ for 12 if the dice is fair.
 - ♦ P_2 would also be different if you used an unfair dice instead of a fair dice--one that produced a higher probability for, say, 2, than for the other numbers.
 - ♦ *The range of real possibilities and their probabilities can be quite different if you begin in different circumstances.*
- F. Notice that you can also use conditional probabilities to represent theories about real world

- ♦ A simple theory would say that P_2 the probability of a 1 is $1/6$, of a 2 is $1/6$ for 2, 3 is $1/6$ and of a 6 is $1/6$ if the dice is fair. In this case, the theory is that if a dice is fair the probability function will look a particular way.
- ♦ Given theory "F" a particular probability function is associated with rolling a dice, $P_2 = f(F | S)$, where S is the set of numbers on the die of interest.
- ♦ (For purists, a pure deterministic causal theory assigns $P = 1$ to the event that has to happen and a $P=0$ to all others--only one event is possible!)

G. Notice that you **can also use conditional probabilities to express uncertainty about the real world.**

- ♦ For example, suppose that the dice landed in a place **hidden from view** and you were asked to predict the value of the number on top.
 - ♦ If you believed that the dice was fair you would admit that you do know the value--really cannot know it even though there is a particular value on top, but you could say that the probability of a 1 is $1/6$, of a 2 is $1/6$ for 2, 3 is $1/6$ and of a 6 is $1/6$ if the dice is fair.
 - ♦ If you did not know whether it was a fair or unfair dice you would also be uncertain about the probabilities as well as the number on top.
 - ♦ Given your belief, theory "F." a particular probability function is associated with rolling a dice, $P_2 = f(F | S)$, where S is the set of numbers on the die of interest.
- H. If you repeated the experiment, you could also **test your theory** and determine whether the dice was really fair.
- ♦ If you repeated enough times, you might eventually decide that you had seen two many "2s" for this to be a fair dice!
 - ♦ *Here, it bears noting that statisticians are quite interested in **how one can reject a theory**--again with various levels of uncertainty.*
 - ♦ In a statistical world, you are rarely 100% sure, only more or less certain that one might be correct about one assessment of the probabilities.
 - ♦ (Draw pictures of confidence intervals for point estimate and for a linear function.)
 - ♦ (Although, given the six sided dice with the number 1 ... 6 on them, you can be absolutely certain that you will not roll a 7, because $P(7)=0$.)
 - ♦ A good deal of the work the econometricians do is of this sort, at least in principle.
- I. One can also **attempt to learn** the correct probability function associated with the dice that one if throwing.

- ♦ There are several techniques that could be used.
 - ♦ The simplest technique would be to throw the dice over and over again and create a table or diagram of the results.
 - ♦ By counting the number of times that each number turns up on top, and dividing by the total number of dice throws, you would have a good estimate of the true probability function.
 - ♦ This is sometimes called the "frequentist" assessment of probability.
- J. Elementary statistical theory implies that the more time that you throw the dice, the more accurate is your estimate of that function--other things being equal.
- ♦ Given an "infinite sample" you would actually know the true probability function that describes the numbers on top of the dice after it is thrown.
 - ♦ **Note that this can be considered a process of learning that converges on the truth even though it does not produce certainty!**

II. The Bayesian Representation of Learning

- A. If one can not spend "all day" collecting data, there is another method of statistical learning that is often called Bayes' Law or **Bayesian Updating**
- B. Given that you have observed "x" and you have a theory "p(s)" that predicts the probability of another value or variable, s. "p(s)" is said to be the **prior probability** (its your initial theory).
- ♦ Suppose that you also now know that the probability of x given s, $f(x | s)$ and you know the probability of x is $f(x)$.
- C. Bayes' law says that you should update your theory given what you have observed.
- ♦ $p(s | x) = p(s) p(x | s) / p(x)$
 - ♦ $p(s | x)$ becomes your new "prior" and the next roll of the dice gives you another "x" which can be regarded as new "information" about p(s)
 - ♦ So, you again update your theory about the probability of s
 - ♦ (For a nice proof and explanation of Bayes' law, see Wikipedia.)
 - ♦ So the learning process goes: (1) form a theory (or a prior), (2) observe the world, (3) update your priors based on what is observed (using Bayes' law to determine "posteriors"). The posterior becomes the new "prior" and the process is repeated.
 - ♦ (In the setting where you know nothing about the problem, you are assumed to use a **"diffuse prior"** which is a uniform distribution.

- D. If you repeat this process, it can be shown that, in the limit, the posterior probability function converges to the actual probability function--in the sense that it is the "expected result" of the process of updating.
- E. This representation of learning is widely used in game theory and economic models.
- ♦ An person is said to be in a Bayesian equilibrium when new information no longer affects the posteriors, e.g. when $p(s | x) = p(s)$
- F. The Bayesian idea can be *generalized to include other similar "empirical" learning functions where $p(s) = f(\text{past observations})$*

III. Weaknesses in the Bayesian representation of learning

- A. Although a convenient and powerful mathematical representation of learning, it clearly has many weaknesses--both as a theory of learning and as a representation of learning.
- B. The conceptual limitations of the usual statistical representations of learning are most apparent in cases in which statistical search or signaling models fail to operate.
- ♦ For example, learning is ruled out within a Bayesian framework whenever some dimensions of the prior probability distribution are missing, because the person in question is ignorant of their existence.
 - ♦ Conventional Bayesian updating in such cases does not occur, because *regardless of the next event observed* the posterior on the missing dimensions remains zero. In such cases, knowledge cannot be accumulated by experience of the sort analyzed in statistical models of learning.
- C. To see this, recall that the posterior probability of event s, given that m has occurred, is the probability of s times the probability of observing m, given that s is true, divided by the probability of event m (Hirshleifer and Riley 1992).
- $$P(s | m) = [P(s) F(m | s) / F(m)] \quad (1)$$
- ♦ If the probability of s is initially assigned a value of zero, $P(s) = 0$, whether implicitly or explicitly, the posterior probability will always be zero whatever the actual probabilities of m and m given s may be.
 - ♦ This holds regardless whether P(s) is assumed to be zero or if one is totally ignorant of the existence of s and *so no probability is assigned to s*.
 - ♦ That is to say, Bayesian updating allows refinements of theories (which can generally be represented as conditional probability functions) over events that *are known* to be possible, but *not* over events completely ignored or completely ruled out a priori.

- D. Learning these “missing dimensions” involves a reduction in ignorance that is fundamentally different from Bayesian updating and similar statistical representations of learning. Priors are not updated when ignorance is reduced, but, rather, *new priors are created* for previously unrecognized possibilities.¹
- E. It also bears noting that reducing ignorance does not always increase one’s sense of certainty.
- F. This contrasts with most statistical notions about which are based on Shannan and Weaver’s (1949) pioneering work on information theory.
 - ♦ Within the context of most statistical characterizations of learning, additional information tends to reduce the variance of the estimates of relevant model parameters, as in ordinary sampling theory.
 - ♦ In the price search illustration, as is often the case for discrete reductions of ignorance, the discovery of new possibilities simultaneously provides new information and *increases uncertainty*.
 - ♦ The combined mall–discount store price distribution has a larger variance than that of the mall alone. The larger world is often more complex and uncertain than previously appreciated.
- G. Discussion: what about “other” as a residual category: magic and miracles?

IV. Discrete representations of information: ignorance and knowledge

- A. As a consequence it is clear that discrete representations of information are also useful.
- B. These will be taken up in future lectures, but are widely used in game theory to characterize “simultaneous” move games, etc.

¹ Binmore (1992, p. 488) suggests that a complete list of the possible outcomes is not even conceptually possible or at least is beyond the capacity of real decision makers. Frydman (1982) argues that agents cannot generally compute optimal forecasts in competitive markets, because they cannot estimate the forecasts of all other agents.

Cyert and DeGroot (1974) provide one of the earliest analyses of the relationship between Bayesian learning functions and rational expectations. Cyert and DeGroot note many cases in which Bayesian learning converges toward rational expectations and market clearing prices. However, they also note several failures (inconsistencies) in cases where the original model used by individuals (firms) is incorrect.