Econ 517, part I
Fall 2002
Probability and Expectation

* What is Probability?

It's the attachment of weights to uncertain prospects. Here are some examples to dilute the dice-throwing and colored-balls-in-urns examples that fill up the beginning chapters of many introductory probability texts.

Example You're going to walk somewhere. It might rain. You have an umbrella you could take with you. Carrying it will be an annoyance if it doesn't rain, but you'll wish you had it if it does rain. Somehow, you decide, and this involves evaluating whether it *probably* will rain or *probably* won't. Also just how annoying is an unnecessary umbrella and just how unpleasant is it to get wet.

You might look at the sky first, taking the umbrella if it's dark, otherwise not taking it. That's statistical inference.

Example A European call option allows you to buy something, say a traded stock, at a future date t^* for a strike price p^*. The option is available in the market now, at time t, for the price q_t. In deciding whether to buy it, you have to assess how likely it is that the stock price p_{t^*} at the strike date is above the strike price, and if so by how much. A European put allows you instead to sell the underlying asset at the strike date at a preset strike price. Clearly a call is worthless at t^* if the actual price at the strike date is below the strike price, and a put is worthless if the actual price at the strike date is above the strike price.

You might look at p_t. If $p_t << p^*$, you might be willing to buy a call option only at a low q_t. That's statistical inference.

---

* Copyright 2002 by Christopher A. Sims. This document may be reproduced for educational and research purposes, so long as the copies contain this notice and are retained for personal use or distributed free.

Contingent claims to a dollar

By buying and selling combinations of put options on the stock, call options on it, and the stock itself, it turns out that we can arrange any pattern of payouts $\pi(p_{t^*})$ at the strike date, as a function of p_{t^*}, that we like.

The details of how this is done were badly garbled in lecture. A detailed discussion is in a separate note.

As is shown in those separate notes, any pattern of yields π that is a piecewise linear function of p can be matched exactly with combinations of purchases and sales of call options, put options, and the underlying asset. Since piecewise linear functions can approximate well any ordinary functions, we can therefore use market prices of puts, calls and the underlying assets to find the price of any reasonable yield function π.

A security with payout function π will have some current price $Q(\pi;t)$. $Q(1;t) = 1/R_t$ is what is usually called the *risk-free rate*, expressed as a discount factor. In a competitive market, $E(\pi) = Q(\pi;t)/Q(1;t)$ behaves like what we will call an **expectation** operator. That is, it maps a space $\mathcal{S}^*$ of functions π into $\mathbb{R}$ and it satisfies

linearity of $\mathcal{S}^*$:

$$\pi_1, \pi_2 \in \mathcal{S}^*, a, b \in \mathbb{R} \Rightarrow a\pi_1 + b\pi_2 \in \mathcal{S}^*$$

linearity of E:

$$(\forall a, b \in \mathbb{R}, \pi_1, \pi_2 \in \mathcal{S}^*) E(a\pi_1 + b\pi_2) \\ = aE(\pi_1) + bE(\pi_2)$$

positivity: $(\forall p)\, \pi(p) \geq 0 \Rightarrow E(\pi) \geq 0$

continuity:

$$\left(\{\pi_n\}_{n=1}^{\infty} \subset \mathcal{S}^* \text{ and } (\forall p \in \mathcal{S}) \pi_n(p) \underset{n\to\infty}{\downarrow} 0\right) \\ \Rightarrow E\pi_n \to 0$$

sure-thing: $E(1_{S^*}) = 1$.

Failure of the linearity of $\mathcal{S}^*$ implies some restriction on markets: securities that we could in principle construct are not priced. The last condition is an automatic consequence of our definition of E. And failure of any of the remaining three conditions would imply an arbitrage opportunity.

Events and probabilities

With an expectation function E in hand, we can attach expectations to sets. If in $\mathcal{S}^*$ there is a function $1_A(p)$ defined to satisfy

$$1_A(p) = \begin{cases} 1 & p \in A \\ 0 & p \notin A\,, \end{cases}$$

then $Q(1_A;t)$ is the price at time t of a dollar delivered at time t^* if and only if at that time $p_{t^*} \in A$. When normalized by R_t to become $E(1_A)$, it is naturally thought of as a a measure of how likely it is, based on information available to the market at time t, that we will find in fact $p_{t^*} \in A$. We usually write such a measure as $P(A)$, where P is thought of as a function mapping subsets of $\mathcal{S}$, the domain of the functions π in $\mathcal{S}^*$, into the interval $[0,1]$.

Interpreting probabilities

- We derived our "market P" above from completeness and competitiveness of markets, with no reference to whether p_{t^*} is "truly random" or to whether the probabilities we have derived are "correct".
- If you or I think the market probability, assigns far too much weight, e.g., to the event $p_{t^*} > \bar{p}$, as evidenced by a high price for the call option, we may take the short side of call option contracts, expecting we will probably make money on them. This is not an arbitrage opportunity.
- Since p_{t^*} will take on a value just once, we can't think of P in terms of frequencies of occurrence of events.

- There are proofs available that, under various assumptions, rational people should make decisions under uncertainty as if they weighted uncertain future events with a probability measure satisfying the properties we are about to derive. These proofs work off the idea that otherwise in a sense nature, or competitors, can "arbitrage" one's behavior.
- Inference is the process of updating a P based on observed data, and the rules for doing this in a consistent way do not depend on whether P is a market measure, a personal set of beliefs of one individual, or is in some sense physical, based on hypothetical or actual frequencies of events in repeated trials.

- While everyone agrees on the rules for updating probabilities based on observations, not everyone would agree that there is nothing more or less to inference than this.

Properties of probability

If P is derived from an expectation operator that is in turn derived from an asset market pricing function as we have just described, then it clearly satisfies

positivity: $(\forall A \in ?) P(A) \geq 0$

countable additivity:

$$\big((\forall i,j \in \mathbb{Z}^+) A_i \cap A_j = \emptyset\big)$$

$$\Rightarrow P\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \sum_{j=1}^{\infty} P(A_j)$$

normalization: $P(\mathbb{S}) = 1$.

σ-fields, measurable functions

To this point we have not been explicit about which functions π constitute the domain of Q and (therefore) E. Also, we have discussed how to generate $P(A)$ for a subset $A \subset \mathbb{S}$, without saying anything about which subsets A we have in mind. And we claimed with a heuristic argument that we could approximate "any" pattern π of contingent payouts dependent on p_{t^*}.

In actual markets prices move in some minimal increment. It used to be "eighths" on the NYSE, now it's "cents". So $\mathbb{S}$ is made up of a countable number of distinct points. This makes $\mathbb{S}$ what is known as a **discrete probability space**. For any such space, there is no need for separate discussion of which subsets of $\mathbb{S}$ P is defined on. It is OK, and natural, to think of $\mathbb{S}^*$ as including *all* bounded functions on $\mathbb{S}$ and to think of $P(A)$ as being defined for *all* subsets $A \subset \mathbb{S}$.

Complexities arising in $\mathbb{R}^k$

As soon as we think about putting probability on $\mathbb{R}$ (not to mention $\mathbb{R}^k$), however, we run in to problems. We often would like to make the probability of every interval of non-zero length non-zero, and to make the probability of every single point zero. It turns out that it is impossible to make a probability measure behave this way on intervals and points and at the same time to have it give every subset of $\mathbb{R}$ a well-behaved probability. Subsets that have to be left out of the domain of definition of such a P on $\mathbb{R}$, called **non-measurable sets**, are exotic creatures that we do not encounter in econometric practice. It takes pages of math just to describe them.

Nonetheless: **Reasons σ-fields are worth studying**

- The related jargon will turn up in papers you should be able to read.
- Stochastic processes and non-parametric models, both important in practice, require thinking about cases where the points in $\mathcal{S}$ are themselves functions, where related issues are more central.
- To model "information available" at a date or to an agent, the most widely used structure is the notion of a class of verifiable events (sets). We want to be able to put probabilities on such events, and hence require that an "information set" correspond to a σ-field of verifiable events. Thus for information arriving over time, the flow of information is represented as an indexed set of σ-fields, say $\{\mathcal{F}_t\}$.

A **σ-field** $\mathcal{F}$ on $\mathcal{S}$ is a class of subsets of $\mathcal{S}$ satisfying:

1. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
2. $\big((\forall j \in \mathbb{Z}^+) A_j \in \mathcal{F}\big) \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$;
3. $\mathcal{S} \in \mathcal{F}$.

Given any class of sets $\mathcal{G}$, there is a well-defined minimal σ-field containing $\mathcal{G}$. Called the σ-field **generated** by $\mathcal{G}$.

The σ-field generated by the intervals in R, or by the k-dimensional rectangles in $\mathbb{R}^k$, ($\{x \in \mathbb{R}^k \mid a_i < x_i < b_i, i = 1, \ldots, k\}$), are called the **Borel** σ-fields in those spaces, and in econometrics and statistics we always work with P's defined on these σ-fields, which we denote $\mathcal{B}$.

This has a corollary: E operators on $\mathbb{R}^k$ are in practice always defined on sets of **measurable** functions, i.e. functions π for which, for any $a \in \mathbb{R}$, $\{p \in \mathbb{R}^k \mid \pi(p) < a\} \in \mathcal{B}$.