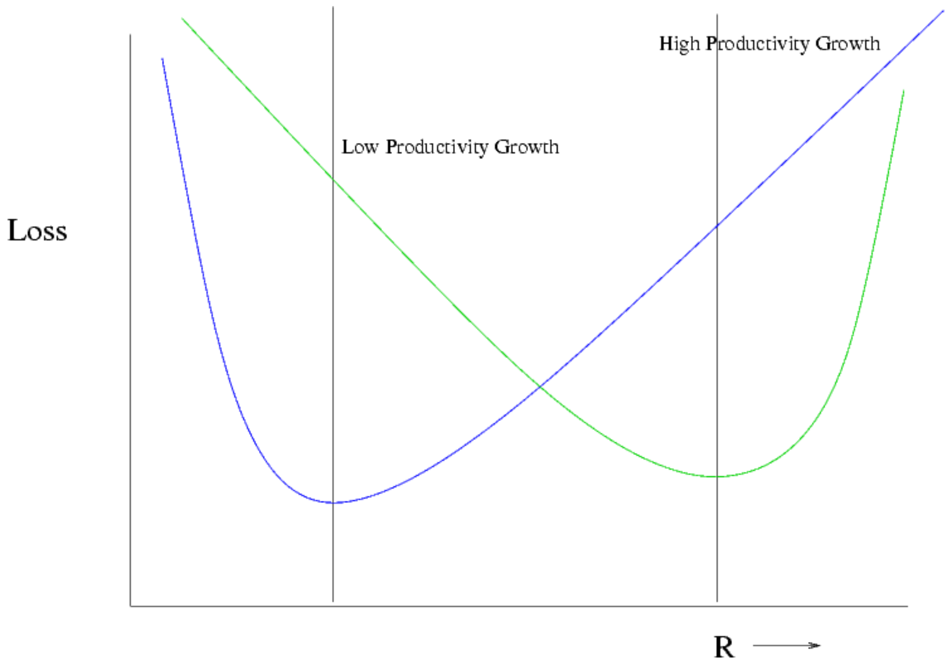

FIGURE 1.

1. MONETARY POLICY WITH UNCERTAIN PRODUCTIVITY GROWTH

In the 90's there was a period when unemployment was falling, to levels that in the past had been associated with accelerating inflation, but there was no sign of rising inflation, and productivity was growing more rapidly than usual. One view about monetary policy at the time was that the high productivity growth was temporary, and that the low unemployment made inflation very likely if monetary policy was not more restrictive. An opposing view was that the high productivity growth was likely to be sustained, that this was making real costs decline, and that inflation was therefore likely to remain modest even at low levels of unemployment.

A stylized view of this situation is represented in Figure 1 The blue line on the graph represents losses as a function of the Federal Funds Rate (the Federal Reserve policy instrument) in case it is true that recent rapid productivity growth will be sustained, while the green line, which reaches its minimum at a higher interest rate, represents losses if the rapid productivity growth is not sustained.

If we are not sure which hypothesis about productivity growth is correct, we can nonetheless (assuming one of them must be correct) be sure that no setting of R to the right of the right-most vertical line, or to the left of the left-hand vertical line passing through the minimum of the blue curve, is a good policy. For each R in these two regions, there is some other R between the two vertical lines that delivers lower loss no matter which hypothesis is correct. The general term for policies like those between the two vertical lines (i.e., undominated policies) is "admissible", with the policies outside the admissible interval termed "inadmissible".

Common sense tells us that we should choose an R nearer the right-hand vertical line if we think productivity growth is likely to drop back to lower levels, while we should pick an R closer to the lower end if we think productivity growth is likely to continue at recent high levels. Furthermore, if we observe new data that tell us something about which hypothesis is true — for example, if we get data on the current quarter's productivity growth rate — we should adjust our beliefs about the relative likelihood of the two hypotheses and adjust our choice of R accordingly.

Decision theory formalizes the process of assessing likelihoods and adjusting decisions in the light of evidence. Statistical inference is the component of the decision process that involves adjusting beliefs in the light of newly observed data.

If the decision process were as simple as this stylized example, there might be little gain from describing it formally with mathematics. The actual monetary policy formation process involves many sources of uncertainty, discussion among many people, and assessment of evidence from many sources. It is therefore worthwhile to have a clear view of the logical framework of the decision process and of how, at least in principle, evidence from all sources and assessments of the consequences of various courses of action ought to interact as decisions are made.

2. PROBABILITIES AND EXPECTATIONS IN A DECISION PROBLEM

Suppose we plot all the pairs of loss function values available on the graph in figure 1. The result would look like figure 2. The lines extending perpendicularly out from the L_p and L_t axes correspond to the two vertical lines in figure 1. The negatively sloped curve connecting them is the set of undominated loss-pairs that can be achieved with various settings of R. Clearly, just as in the similar diagram that arises in intermediate microeconomic-theory discussion of consumer theory, every one of the points on this curve corresponds to a tangent line, like the one that has been drawn in on the figure. Such a line will be a linear function, of the general form shown on the figure — $p_pL_p + p_tL_t = A$. The line will be unaffected if we multiply this equation by any non-zero constant, so we might as well normalize by requiring that $p_p + p_t = 1$. The point at which the line touches the curve clearly is the point that minimizes $p_pL_p + p_tL_t$ in the set of available loss pairs. p_p and p_t are **probabilities** on the "persistent" and "temporary" hypotheses, and the value A on the right-hand side of the equation is the **expectation** of losses for that choice of interest rate and the p_p, p_t probabilities.

It is clear in this simple example that every choice of R that is not dominated can be described as minimizing expected loss for some choice of probability weights. We might choose an R that delivers a point on the losses curve without bothering to compute the corresponding p's and expected losses. On the other hand, even if we had done so, it might be useful to compute the p's and the A (expected loss) as a check on whether our choice was actually reasonable.

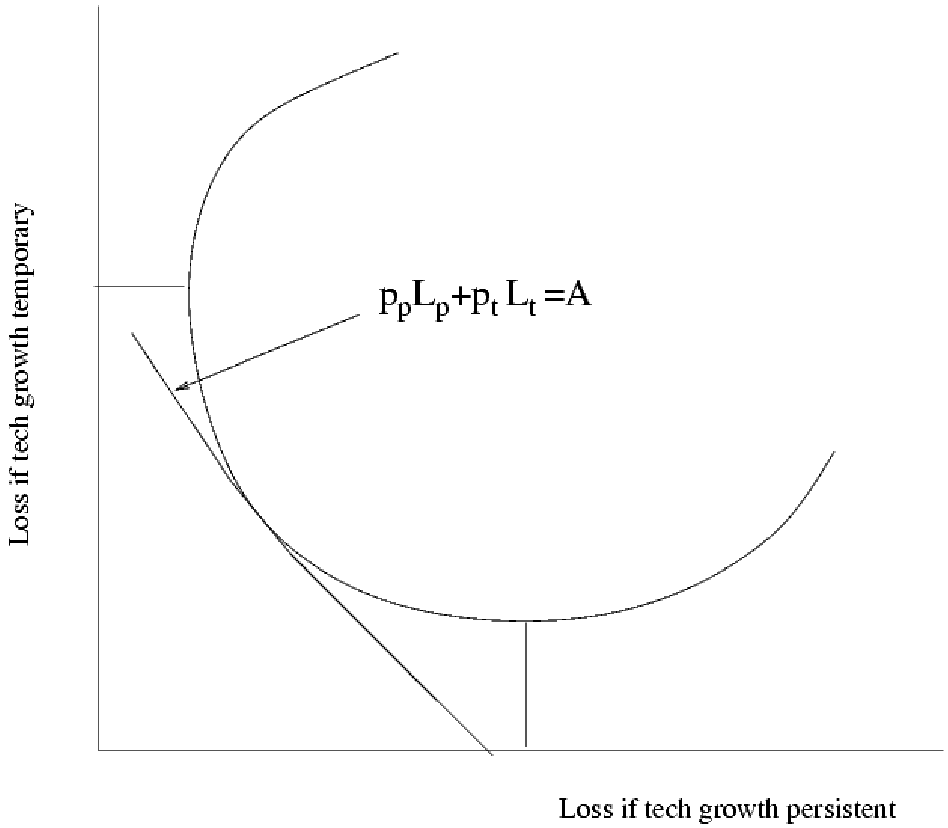

FIGURE 2.

The result here, that admissible decisions can be represented as minimizing expected losses under some set of probability weights, is true under quite general conditions. It underlies one main interpretation of probability theory, that probabilities are weights on uncertain prospects that underly optimal decisions.

Note that we have not introduced any notion of repeated trials or of frequencies of occurrence in arriving at this interpretation of probability.

3. PROBABILITY FROM ARBITRAGE-FREE PRICING IN COMPETITIVE MARKETS

Suppose we have a set S of possible states of the world, or contingencies, and we have a set of N securities with yield functions $y_j : S \to \mathbb{R}$, $j = 1, \ldots, \infty$. A yield function describes the dollar return, $y_j(\omega)$ that security j delivers when the state of the world turns out to be $\omega \in S$. We don't know what the state of the world will turn out to be, so we don't know what the yields on the securities will be. Now suppose these securities are sold on a competitive market, where each market participant sees the opportunity to buy or sell arbitrary amounts of any of the securities at the market price. In fact, we will assume that any participant can create new securities whose yields are of the form $z = \sum a_j y_j$, where the a_j are arbitrary positive real numbers, and sell or buy the new securities at market prices. Let P be the function that maps a security's yield function z into its market price $Q(z)$.

If this market is to function as we have hypothesized, it must not allow arbitrage. This requires in turn that it must be linear, i.e. that if $z = \sum a_j y_j$, $Q(z) = \sum a_j Q(y_j)$. If instead we had $Q(z) > \sum a_j Q(y_j)$, someone could buy ka_j units of each of the N securities and sell k units of the z security, paying the yield on the z security from the earnings on the y_j securities. The profits from this transaction could be made arbitrarily high by picking k arbitrarily large. A similar arbitrage opportunity would arise if $Q(z) < \sum a_j Q(a_j)$, except to exploit this one would buy the z security and sell the a_j securities. We also require that, so long as $z(\omega) > 0$ for all $\omega \in S$, $Q(z) > 0$. Otherwise it is possible at no cost to guarantee a positive return, so again profits can be made arbitrarily high.

Finally, suppose it is possible to construct, from linear combinations of the y_j's, a security whose yield is constant across S — that is, a risk-free security. Then since we can scale securities any way we like and still price them, we can construct a security with yield $x(\omega) \equiv 1$, and set $\Phi = Q(x)$. Φ is what is usually called the risk-free discount factor, and Φ^{-1} the risk-free gross interest rate. The function defined by $E[y] = Q(y)/\Phi$ then has most of the properties of mathematical expectation. If S is a finite space with M elements, and if $N \geq M$, with the y_j functions linearly independent, we can price, for each $\omega_i \in S$, a security with yield e_i defined as $e_i(\omega_i) = 1$, $e_i(\omega_j) = 0$, $j \neq i$. If we set $p(\omega_j) = E[e_i]$, the N $p(\omega_j)$'s will be non-negative and sum to one, and E will be the expectation operator with respect to the probability defined by these weights. The probability we have arrived at this way is what as known in the finance literature as the **market measure** or **market probability**, and the E operator is called the **market expectation**.

This kind of probability arises not only with no connection to frequency or repetition but also with no reference to loss functions or decision theory. The market probability corresponds to the beliefs of no individual participating in the market and can be deduced from market data. In that sense it is "objective". But it will ordinarily not have the properties associated with probabilities based on frequencies. For example, it will generally be true that, if markets are observed repeatedly, the average observed value of $z(\omega) - E[z]$ does not tend to zero as the number of repetitions we observe increases.

4. Physical probability

Decision-theoretic probability and market probability both arise from beliefs, rather than being a measurable physical notion. There are two approaches to interpreting probability as a property of the physical world, as opposed to arising from people's beliefs.

One is to derive probability from considerations of symmetry. One may examine a coin carefully, weighing it, balancing it, etc., and expect to reach a conclusion as to whether it is "fair" — that is, equally likely to come up heads or tails when flipped properly. One can similarly check a standard six-sided die to verify that it is symmetric, and thus equally likely to land with any one of its six sides up.

The fairness of the coin or the die is then a physical fact about which there should not be differing beliefs (once the measurements are done). Decision-makers whose losses depend on the outcome of a toss of this coin or die should weight each side of the die or coin equally in forming probabilities and expectations. The market should value securities that pay y_h when heads occurs and y_t when tails occurs exactly as it does securities that pay y_h when tails occurs and y_t when heads occurs.

If we can find a collection of non-overlapping subsets of S to which we give equal probability, we can then construct probabilities for more complicated sets that we can build from these elementary ones — e.g. the probability of a die coming up with a prime number on its face.

Another approach is to consider situations where there are many replicas of the state space S, indexed by $t = 1,\dots,\infty$ and where we will observe for each t the value of a function z_t defined onS_t. If we assume that, for some collection of functions $\{f_j\}$, $T^{-1}\sum_1^T f_j(z_t)$ always converges to a limit as $T \to \infty$, we can treat the mapping from f_j to this limit as an expectation operator $E[f_j]$. It is not hard to see that this version of E must have the same properties we derived for market expectation in section 3, and therefore that we can generate probabilities from these averages as in that section.

Few would disagree with the idea that when the conditions allowing building probability from physical symmetry coniderations or from limiting frequencies are met, probabilities should be built that way. Furthermore, in characterizing scientific results it makes sense to maintain a clean distinction between such physically based probabilities and probabilities that do not have such a foundation. On the other hand, in most real-world decision problems most of the uncertainty has to be given weights without any possibility of appeal to such long-run frequencies or physical symmetry.

Some descriptions of the foundations of probability theory seem to imply that the interpretations we have given here are in conflict, as if they are mutually exclusive. In fact, there is no conflict at all between decision-theoretic interpretations of probability and the physical ones. Physical probabilities are, from the decision-theoretic perspective, a special case. Conflict only arises when physical probabilities are claimed to be the only legitimate type of probability.

Market probabilities are also understandable from a decision-theoretic perspective. In a competitive market with rational agents behaving according to the postulates of decision theory, it is possible to derive the form of the market probabilities from knowledge of the probabilities used by the individuals participating in the market (plus knowledge of their budget constraints and utility functions).

5. Defining properties of expectations and probabilities

We can start from expectations and derive probabilities, or vice versa. In either case we begin with the space S of states of the world. For expectations, we then introduce a space F of functions $f : S \to W$, where W is a linear space. Functions defined on S are called **random variables**. [1] The defining properties of an expectation operator $E : F \to W$ are

[1]Often, though, the term "random variable" is reserved for the case $W = \mathbb{R}$. When $W = \mathbb{R}^n$, f is often called a "random vector" or "vector of random variables"; when $W = \mathbb{R}^{\mathbb{Z}^+}$ (the space of sequences of real

(1) Ef is defined for all $f \in F$ and F is a linear space.
(2) If f and g are each in F and a and b are real numbers, $E[af+bg] = aEf + bEg$.
(3) If $f(\omega) \geq 0$ for all $\omega \in S$, $Ef \geq 0$.
(4) If $\forall(\omega \in S) f(\omega) = c$, $Ef = c$. (Sometimes this is written loosely as "$E[c] = c$".)
(5) If $f_n \in F$ for every n, and if $f_n(\omega) \to 0$ monotonically as $n \to \infty$ for each $\omega \in S$, $E[f_n] \to 0$ as $n \to \infty$.

In order to connect expectation to probability we need one more condition, which we state here for the case where $W = \mathbb{R}$:

(6) For any $f \in F$ the function f^+, defined by

$$f^+(\omega) = \begin{cases} f(\omega) & \text{if } f(\omega) > 0 \\ 0 & \text{if } f(\omega) \leq 0 \end{cases},$$

is also in F.

This last condition and (5) are not needed if we restrict ourselves to S with finitely many elements.

To start with probability, we begin with a family $\{\mathcal{F}\}$ of subsets of S rather than a family of functions on S. We require that $\mathcal{F}$ be a σ-field, which means

(1) If A_i is in $\mathcal{F}$ for every $i = 1, \ldots, \infty$, then $\bigcup_i A_i \in \mathcal{F}$.
(2) $S \in \mathcal{F}$.
(3) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$. (A^c is the complement of A in S.)

Then a probability measure on $\mathcal{F}$ is a function $P : \mathcal{F} \to [0,1]$ satisfying

(1) For any $A \in \mathcal{F}$, $P[A] \geq 0$;
(2) $P[S] = 1$;
(3) For any disjoint sets $A, B \in \mathcal{F}$, $P[A \cup B] = P[A] + P[B]$; ($A$ and B disjoint means $A \cap B = \emptyset$.)
(4) If $A_j \in \mathcal{F}$ for every integer j, then

$$P\left[\bigcup_{j=1}^{n} A_j\right] \xrightarrow[n\to\infty]{} P\left[\bigcup_{j=1}^{\infty} A_j\right].$$

We say a collection $\mathcal{G}$ of sets **generates** the σ-field $\mathcal{F}$ if $\mathcal{F}$ is the smallest σ-field containing $\mathcal{G}$. The probability P corresponding to a given expectation operator E can be found as follows. Let $\mathcal{G}$ be the collection of all subsets of S of the form $\{\omega \in S \mid f(\omega) < a)\}$ for some $f \in F$ and some real number a. Let $\mathcal{F}$ be the σ-field generated by $\mathcal{G}$ and define the probability of a set $A \in \mathcal{F}$ to be $P[A] = E[\mathbf{1}_A]$, where $\mathbf{1}_A$ is the indicator function for the set A. It can be shown that the assumptions we have made on E assure that indicator functions for these sets are in F. It will then be true that we can write

$$E[f] = \int_S f(\omega)\, dP(\omega). \tag{1}$$

numbers), f is usually called a "discrete-time stochastic process"; and when $W = \mathbb{R}^{\mathbb{R}^+}$, f is usually called a "continuous-time stochastic process".

That is, the expectation of a random variable is the same thing as its integral with respect to the associated probability. The integral appearing in (1) is not a standard Calculus I (Riemann) integral, however. We will discuss the general interpretation of this integral later. For now, it is enough to note that when S has a finite or countable number of points, the j'th point in S has a probability p_j and

$$\int_S f(\omega)\, dP(\omega) = \sum_{\omega_j \in S} p_j f(\omega_j)\,, \tag{2}$$

while if S is $\mathbb{R}^n$, the most common special case is one in which the probability has a **density function** p and we can write

$$\int_S f(\omega)\, dP(\omega) = \int\int \ldots \int p(\omega) f(\omega)\, d\omega_1\, d\omega_2 \ldots d\omega_n\,, \tag{3}$$

where what appears on the right *is* an ordinary Riemann integral.

It should be clear how we go from a probability to a corresponding E operator: find a linear space of f's for which the integral on the right of (1) is defined, and define E by (1).