mathematical appendix

This appendix has three purposes: to remind some readers of the definitions of terms they have seen before, to give other readers an idea of what the terms mean, and to list a few theorems for reference. In accordance with these limited purposes, some terms such as "boundary point" are left undefined. For fuller exposition, see Rudin (1964) on real analysis, Debreu's *Theory of Value* (1959), and Chiang (1984) and Takayama (1985) on mathematics for economists. Intriligator (1971) and Varian (1992) both have good mathematical appendices and are strong in discussing optimization, and Kamien & Schwartz (1991) covers maximizing by choice of functions. Border's 1985 book is entirely about fixed point theorems. Stokey & Lucas (1989) is about dynamic programming. Fudenberg & Tirole (1991a) is the best source of mathematical theorems for use in game theory.

The web is very useful for mathematical definitions. See http://en.wikipedia.org, http://mathworld.wolfram.com, and http://planetmath.org.

*A.1 Notation

$\sum$ **Summation.** $\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3$.

Π **Product.** $\Pi_{i=1}^{3} x_i = x_1 x_2 x_3$.

$|x|$ **Absolute value** of x. If $x \geq 0$ then $|x| = x$ and if $x < 0$ then $|x| = -x$.

$|$ **"Such that,"** "given that," or "conditional upon." $\{x | x < 3\}$ denotes the set of real numbers less than three. $Prob(x | y < 5)$ denotes the probability of x given that y is less than 5.

$:$ **"Such that."** $\{x : x < 3\}$ denotes the set of real numbers less than three. The colon is a synonym for $|$.

$\mathbf{R}^n$ The set of n-dimensional vectors of **real numbers** (integers, fractions, and the least upper bounds of any subsets thereof).

$\{x, y, z\}$ **A set of elements** x, y, and z. The set $\{3, 5\}$ consists of two elements, 3 and 5.

$\in$ **"Is an element of."** $a \in \{2, 5\}$ means that a takes either the value 2 or 5.

$\subset$ **Set inclusion.** If $X = \{2, 3, 4\}$ and $Y = \{2, 4\}$, then $Y \subset X$ because Y is a subset of X.

$[x, y]$ The **closed interval** with endpoints x and y. The interval $[0, 1{,}000]$ is the set $\{x | 0 \leq x \leq 1{,}000\}$. Square brackets are also used as delimiters.

(x, y) The **open interval** with endpoints x and y. The interval $(0, 1{,}000)$ is the set $\{x | 0 < x < 1{,}000\}$. $(0, 1{,}000]$ would be a half-open interval, the set $\{x | 0 < x \leq 1{,}000\}$. Parentheses are also used as delimiters.

$x!$ **x-factorial.** $x! = x(x - 1)(x - 2) \ldots (2)(1)$. $4! = 4(3)(2)(1) = 24$.

$\binom{a}{b}$ The number of unordered combinations of b elements from a set with a elements. $\binom{a}{b} = a!/b!(a - b)!$, so $\binom{4}{3} = 4!/3!(4 - 3)! = 24/6 = 4$. (See *combination* and *permutation* below.)

$\times$ The **Cartesian product.** $X \times Y$ is the set of points $\{x, y\}$, where $x \in X$ and $y \in Y$.

ϵ An **arbitrarily small positive number.** If my payoff from both *Left* and *Right* equals 10, I am indifferent between them; if my payoff from *Left* is changed to $(10 + \epsilon)$, I prefer *Left*.

$\sim$ We say that $X \sim F$ if the random variable X is **distributed according to** distribution F.

$\exists$ **"There exists. . ."** $\exists x > 0\colon 9 - x^2 = 0$.

$\forall$ **"For all. . ."** $\forall x \in [0, 3]$, $x^2 < 10$.

$\equiv$ **"Equals by definition"** "For clarity, let us define the average income $x \equiv \theta^{w-1}/(a - 1)^2 + b^2 + c$ for use in the expressions below."

$\rightarrow$ If f **maps** space X into space Y then $f\colon X \rightarrow Y$.

$\dfrac{df}{dx}, \dfrac{d^2f}{dx^2}$ The **first and second derivatives** of a function. If $f(x) = x^2$ then $df/dx = 2x$ and $d^2f/dx^2 = 2$.

f', f'' The **first and second derivatives** of a function. If $f(x) = x^2$ then $f' = 2x$ and $f'' = 2$. Primes are also used on variables (not functions) for other purposes: x' and x'' might denote two particular values of x.

$\dfrac{\partial f}{\partial x}, \dfrac{\partial^2 f}{\partial x \partial y}$ **Partial derivatives** of a function. If $f(x, y) = x^2 y$ then $\partial f/\partial x = 2xy$ and $\partial^2 f/\partial x \partial y = 2x$.

y_{-i} The set y minus element i. If $y = \{y_1, y_2, y_3\}$, then $y_{-2} = \{y_1, y_3\}$.

$Max(x, y)$ The **maximum** of two numbers x and y. $Max(8, 24) = 24$.

$Min(x, y)$ The **minimum** of two numbers x and y. $Min(5, 3) = 3$.

$\lceil x \rceil$ **Ceiling**(x). A number rounded up to the nearest integer. $\lceil 4.2 \rceil = 5$. This notation is not well known in economics.

$\lfloor x \rfloor$ **Floor**(x). A number rounded down to the nearest integer. $\lfloor 6.9 \rfloor = 6$. This notation is not well known in economics.

Sup X   The **supremum (least upper bound)** of set X. If $X = \{x | 0 \leq x < 1,000\}$, then *sup X* $= 1,000$. The supremum is useful because sometimes, as here, no maximum exists.

Inf X   The **infimum (greatest lower bound)** of set X. If $X = \{x | 0 \leq x < 1,000\}$, then *inf X* $= 0$.

Argmax   The **argument that maximizes** a function. If $e^* = argmax\ EU(e)$, then e^* is the value of e that maximizes the function $EU(e)$. The argmax of $f(x) = x - x^2$ is $1/2$.

Maximum   The **greatest value** that a function can take. $Maximum(x - x^2) = 1/4$ for $x \geq 0$ (at $x = 0$).

Minimum   The **least value** that a function can take. $Minimum(-5 + x^2) = -5$ (at $x = 0$).

*A.2   The Greek Alphabet

A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ	delta	Π	π	pi
E	ϵ or ε	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ	sigma
H	η	eta	T	τ	tau
Θ	θ	theta	Υ	υ	upsilon
I	ι	iota	Φ	ϕ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

*A.3   Glossary

almost always See "generically."

annuity A riskless security paying a constant amount each year for a given period of years, with the amount conventionally paid at the end of each year.

closed A closed set in $\mathbf{R}^n$ includes its boundary points. The set $\{x: 0 \leq x \leq 1,000\}$ is closed.

combination The number of unordered sets of b elements from a set with a elements, denoted $\binom{a}{b} = a!/b!(a-b)!$. If we form sets of 2 elements from the set $A = \{w, x, y, z\}$, the possibilities are $\{w, x\}, \{w, y\}, \{w, z\}, \{x, y\}, \{x, z\}, \{y, z\}$. Thus, $\binom{4}{2} = 4!/2!(4-2)! = 24/6 = 6$. (See *permutation* for the ordered version.)

compact If set X in $\mathbf{R}^n$ is closed and bounded, then X is compact. Outside of Euclidean space, however, a set being closed and bounded does not guarantee compactness.

complete metric space A metric space that includes the limits of all possible Cauchy sequences. All compact metric spaces and all Euclidean spaces are complete.
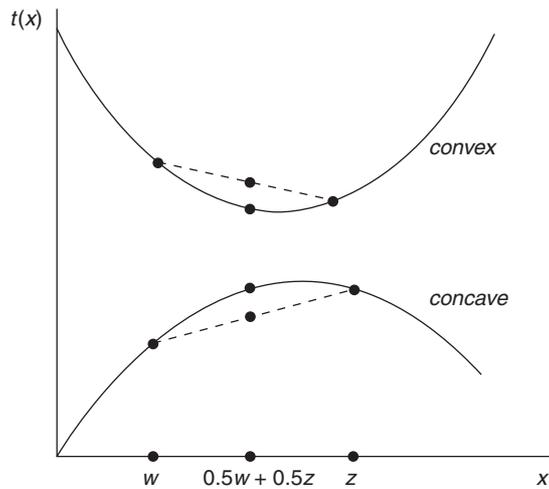
Figure A.1 Concavity and convexity.

concave function The continuous function $f(x)$ defined on interval X is concave if for all elements w and z of X, $f(0.5w + 0.5z) \geq 0.5f(w) + 0.5f(z)$. If f maps **R** into **R** and f is concave, then $f'' \leq 0$. See figure A.1.

continuous function Let $d(x, y)$ represent the distance between points x and y. The function f is continuous if for every $\epsilon > 0$ there exists a $\delta(\epsilon) > 0$ such that $d(x, y) < \delta(\epsilon)$ implies $d(f(x), f(y)) < \epsilon$.

continuum A continuum is a closed interval of the real line, or a set that can be mapped one-to-one onto such an interval.

contraction The mapping $f(x)$ is said to be a contraction if there exists a number $c < 1$ such that for the metric d of the space X,

$$d(f(x), f(y)) \leq c * d(x, y), \quad \text{for all } x, y \in X. \tag{A.1}$$

convex function The continuous function $f(x)$ is convex if for all elements w and z of X, $f(0.5w + 0.5z) \leq 0.5f(w) + 0.5f(z)$. See figure A.1. Convex functions are only loosely related to convex sets.

convex set If set X is convex, then if you take any two of its elements w and z and a real number t: $0 \leq t \leq 1$, then $tw + (1 - t)z$ is also in X.

correspondence A correspondence is a mapping that maps each point to one or more other points, as opposed to a function, which only maps to one.

domain The domain of a mapping is the set of elements it maps from – something like the land that it can alter as it pleases. (The mapping maps from the domain onto the range.)

function If f maps each point in X to exactly one point in Y, f is called a function. The two mappings in figure A.1 are functions, but the mapping in figure A.2 is not.

generically If a fact is true on set X generically, "except on a set of measure zero," or "almost always," then it is false only on a subset of points Z that have the property that if a point is randomly chosen using a density function with support X, a point in Z is chosen with probability zero. This implies that if the fact is false on $z \in \mathbf{R}^n$ and z is perturbed by adding a random amount ϵ, the fact is true on $(z + \epsilon)$ with probability one.
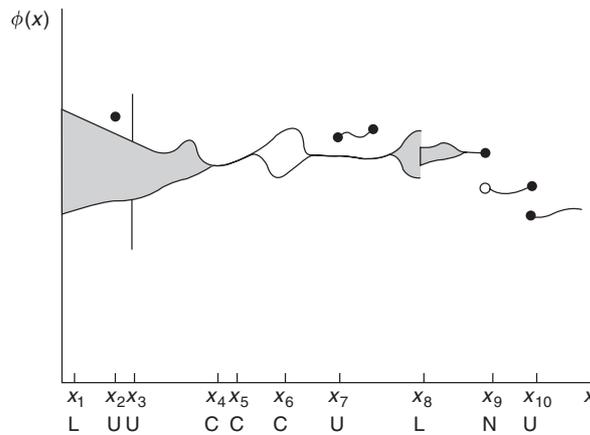
Note: Points at which the correspondence is just upper semi continuous are labelled *U*; just lower semicontinuous, *L*; both, *C*; and neither, *N*.

Figure A.2 Upper semicontinuity.

integration by parts This is a technique to rearrange integrals so they can be solved more easily. It uses the formula

$$\int_{z=a}^{b} g(z)h'(z)dz = g(z)h(z)\Big|_{z=a}^{b} - \int_{z=a}^{b} h(z)g'(z)dz. \tag{A.2}$$

To derive this, differentiate $g(z)h(z)$ using the chain rule, integrate each side of the equation, and rearrange.

Lagrange multiplier The Lagrange multiplier λ is the marginal value of relaxing a constraint in an optimization problem. If the problem is

$$\left\{ \underset{x}{Maximize} \ x^2 \text{ subject to } x \leq 5 \right\} \text{ then } \lambda = 2x^* = 10.$$

lattice A lattice is a partially ordered set (the $\geq$ ordering is defined) where for any two elements a and b, the values $inf(a, b)$ and $sup(a, b)$ are also in the set. A lattice is **complete** if the infimum and supremum of each of its subsets are in the lattice.

Leibniz's integral rule This is the rule for differentiation under the integral sign.

$$\frac{\partial}{\partial z} \int_{a(z)}^{b(z)} f(x, z)dx = f(b(z), z)\frac{\partial b(z)}{\partial z} - f(a(z), z)\frac{\partial a(z)}{\partial z} + \int_{a(z)}^{b(z)} \frac{\partial f(x, z)}{\partial z}dx. \tag{A.3}$$

list (or *n*-Tuple) A set of *n* elements in which the element positions are ordered. $(Up, Down, Up)$ is a list, but $(Down, Up, Up)$ and $(Up, Down)$ are different lists. See *set* and *multiset*.

lower semicontinuous correspondence The correspondence ϕ is lower semicontinuous at the point x_0 if

$$x_n \rightarrow x_0, \ y_0 \in \phi(x_0), \quad \text{implies } \exists y_n \in \phi(x_n) \ \text{such that} \ y_n \rightarrow y_0, \tag{A.4}$$

which means that associated with every x sequence leading to x_0 is a y sequence leading to its image. See figure A.2. This idea is not as important as upper semicontinuity.

maximand A maximand is what is being maximized. In the problem "Maximize $f(x, \theta)$ by choice of x," the maximand is f.

mean-preserving spread See the **Risk** section below.

measure zero See "generically."

metric The function $d(w, z)$ defined over elements of set X is a metric if (1) $d(w, z) > 0$ if $w \neq z$ and $d(w, z) = 0$ if and only if $w = z$; (2) $d(w, z) = d(z, w)$; and (3) $d(w, z) \leq d(w, y) + d(y, z)$ for points $w, y, z \in X$.

metric space Set X is a metric space if it is associated with a metric that defines the distance between any two of its elements.

multiset A set of n elements in which position in the listing is does not matter but multiplicity does. $(Up, Down, Up)$ is a multiset, and $(Down, Up, Up)$ is the same multiset, but $(Up, Down)$ is a different multiset. See *list* and *set*.

n-Tuple (or Tuple). See *list*.

one-to-one The mapping $f: X \rightarrow Y$ is one-to-one if every point in set X maps to a different point in Y, so $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$. An example is $f(x) = x/2$ with $X = [0, 1]$ and $Y = [0, 2]$.

onto The mapping $f: X \rightarrow Y$ is onto Y if every point in Y is mapped onto by some point in X. An example is $f(x) = x^2$ with $X = [-1, 1]$ and $Y = [0, 1]$.

open In the space $\mathbf{R}^n$, an open set is one that does not include all its boundary points. The set $\{x: 0 \leq x < 1{,}000\}$ is open (even though it does include *one* of its boundary points). In more general spaces, an open set is a member of a topology.

permutation The number of lists (sets with an order) of b elements from a set with a elements, which equals $a!/(a - b)!$ If we form sets of 2 elements from the set $A = \{w, x, y, z\}$, the possibilities are

$$\{w, x\}, \{x, w\}, \{w, y\}, \{y, w\}, \{w, z\}, \{z, w\}, \{x, y\}, \{y, x\}, \{x, z\}, \{z, x\}, \{y, z\}, \{z, y\}.$$

The number of these is $4!/(4 - 2)! = 24/2 = 12$. (See *combination* for the unordered version.)

perpetuity A riskless security paying a constant amount each year in perpetuity, with the amount conventionally paid at the end of each year.

quasi-concave The continuous function f is quasi-concave if for $w \neq z$, $f(0.5w + 0.5z) > min[f(w), f(z)]$, or, equivalently, if the set $\{x \in X | f(x) > b\}$ is convex for any number b. Every concave function is quasi-concave, but not every quasi-concave function is concave.

quasilinear utility A utility function is quasilinear in variable w if a monotonic transformation can make it linear in w and w is separable from all other variables in the utility function. $u(w, x) = w + \sqrt{x}$ and $u(w, x) = log(w + \sqrt{x})$ are quasilinear; $u(w, x) = wx$ and $u(w, x) = log(w) + \sqrt{x}$ are not.

range The range of a mapping is the set of elements to which it maps – the property over which it can spew its output. (The mapping maps from the domain onto the range.)

risk See the **Risk** section below.

set A set is a collection of objects in which position of listing and multiplicity do not matter. $\{Up, Down\}$, $\{Down, Up\}$, and $\{Up, Down, Down\}$ are all the same set of 2 elements. See *list* and *multiset*.

stochastic dominance See the **Risk** section below.

strict The word "strict" is used in a variety of contexts to mean that a relationship does not hold with equality or is not arbitrarily close to being violated. If function f is concave and $f' > 0$, then $f'' \leq 0$, but if f is strictly concave, then $f'' < 0$. The opposite of "strictly" is "weakly." The word "strong" is used as a synonym for "strict."

supermodular See the **Supermodularity** section below.

support The support of a probability distribution $F(x)$ is the closure of the set of values of x such that the density is positive. If each output between 0 and 20 has a positive probability density, and no other output does, then the support of the output distribution is [0, 20].

topology Besides denoting a field of mathematics, a topology is a collection of subsets of a space called "open sets" that includes (1) the entire space and the empty set, (2) the intersection of any finite number of open sets, and (3) the union of any number of open sets. In a metric space, the metric "induces" a topology by defining an open set. Imposing a topology on a space is something like defining which elements are close to each other, which is easy to do for $\mathbf{R}^n$, but not for every space (e.g., spaces consisting of functions or of game trees).

upper semicontinuous correspondence The correspondence $\phi: X \rightarrow Y$ is upper semicontinuous at point x_0 if

$$x_n \rightarrow x_0, \ y_n \in \phi(x_n), \ y_n \rightarrow y_0, \quad \text{implies} \ y_0 \in \phi(x_0), \tag{A.5}$$

which means that every sequence of points in $\phi(x)$ leads to a point also in $\phi(x)$. See figure A.2. An alternative definition, appropriate only if Y is compact, is that ϕ is upper semicontinuous if the set of points $\{x, \phi(x)\}$ is closed.

vector A vector is a list (a set with order) that has a certain structure I will not describe here, but which, for example, a list of real numbers, a point in $\mathbf{R}^n$, will satisfy. The point $(2.5, 3, -4)$ is a vector in $\mathbf{R}^3$.

weak The word "weak" is used in a variety of contexts to mean that a relationship might hold with equality or be on a borderline. If f is concave and $f' > 0$, then $f'' \leq 0$, but to say that f is weakly concave, while technically adding nothing to the meaning, emphasizes that $f'' = 0$ under some or all parameters. The opposite of "weak" is "strict" or "strong."

*A.4   Formulas and Functions

$log(xy) = log(x) + log(y)$.

$log(x^2) = 2log(x)$.

$a^x = (e^{log(a)})^x$.

$e^{rt} = (e^r)^t$.

$e^{a+b} = e^a e^b$.

$a > b \Rightarrow ka < kb, \quad \text{if } k < 0$.

The Quadratic Formula: Let $ax^2 + bx + c = 0$. Then $x = -b \pm \sqrt{b^2 - 4ac}/2a$.

Derivatives

$f(x)$	$f'(x)$
x^a	ax^{a-1}
$1/x$	$-\dfrac{1}{x^2}$
$\dfrac{1}{x^2}$	$-\dfrac{2}{x^3}$
e^x	e^x
e^{rx}	re^{rx}
$log(x)$	$\dfrac{1}{x}$
$log(ax)$	$\dfrac{1}{ax}a = \dfrac{1}{x}$
a^x	$a^x log(a)$
$f(g(x))$	$f'(g(x))g'(x)$

See also Lawrence Spector's 2006 "An Approach to Calculus," http://www.themath-page.com/aCalc/exponential.htm

Table A.1   Some useful functional forms

$f(x)$	$f'(x)$	$f''(x)$	*Slope for $x > 0$*	*Curvature*
$log(x)$	$\dfrac{1}{x}$	$-\dfrac{1}{x^2}$	Increasing	Concave
$\sqrt{x}$	$\dfrac{1}{2\sqrt{x}}$	$-\dfrac{1}{4x^{(3/2)}}$	Increasing	Concave
x^2	$2x$	2	Increasing	Convex
$\dfrac{1}{x}$	$-\dfrac{1}{x^2}$	$\dfrac{2}{x^3}$	Decreasing	Convex
$7 - x^2$	$-2x$	-2	Decreasing	Concave
$7x - x^2$	$7 - 2x$	-2	Increasing/Decreasing	Concave

The signs of derivatives can be confusing. The function $f(x) = x^2$ is increasing at an *increasing* rate, but the function $f(x) = 1/x$ is decreasing at a *decreasing* rate, even though $f'' > 0$ in each case.

Determinants

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}.$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} \\ - a_{12}a_{21}a_{33} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$

*A.5 Probability Distributions

The definitive listing of probability distributions and their characteristics is the three-volume series of Johnson & Kotz (1970). A few major distributions are listed here. A **probability distribution** is the same as a **cumulative density function** for a continuous distribution. Any single value has infinitesimal probability if the distribution is continuous (unless there is a probability atom there), so rather than speaking of the probability of a value, we speak of the probability of the value being in an interval, or of the **density** at a single value.

The Exponential Distribution

The exponential distribution, which has the set of nonnegative real numbers as its support, has the density function for mean λ of

$$f(x) = \frac{e^{-x/\lambda}}{\lambda}. \tag{A.6}$$

The cumulative density function is

$$F(x) = 1 - e^{-x/\lambda}. \tag{A.7}$$

The Uniform Distribution

A variable is uniformly distributed over support X if each point in X has equal probability. If the support is $[\alpha, \beta]$, the mean is $(\alpha + \beta)/2$, the density is

$$f(x) = \begin{cases} 0 & x < \alpha, \\ \dfrac{1}{\beta - \alpha} & \alpha \le x \le \beta, \\ 0 & x > \beta, \end{cases} \tag{A.8}$$

and the cumulative density function is

$$F(x) = \begin{cases} 0 & x < \alpha, \\ \dfrac{x - \alpha}{\beta - \alpha} & \alpha \le x \le \beta, \\ 1 & x > \beta. \end{cases} \tag{A.9}$$

The Normal Distribution

The normal distribution is a two-parameter single-peaked distribution which has as its support the entire real line. The density function for mean μ and variance σ^2 is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \tag{A.10}$$

The cumulative density function is the the integral of this, often denoted $\Phi(x)$, which cannot be simplified analytically. You can find values on web sites such as http://www.math2.org/math/stat/distributions/z-dist.htm or from a spreadsheet program.

The Lognormal Distribution

If $log(x)$ has a normal distribution, x has a lognormal distribution. This is a skewed distribution which has the set of positive real numbers as its support since the logarithm of a negative number is not defined. The mean is $e^{\sigma^2/2}$.

A.6   Supermodularity

Suppose that there are N players in a game, subscripted by i and j, and that player i has a strategy consisting of $\bar{s}^i$ elements, subscripted by s and t, so his strategy is the vector $y^i = (y^i_1, \ldots, y^i_{\bar{s}^i})$. Let his strategy set be S^i and his payoff function be $\pi^i(y^i, y^{-i}; z)$, where z represents a fixed parameter. We say that the game is a **supermodular game** if the following four conditions are satisfied for every player $i = 1, \ldots, N$:

(A1) S^i is a complete lattice.

(A2) $\pi^i : S \to R \cup \{-\infty\}$ is order semicontinuous in y^i for fixed y^{-i}, and order continuous in y^{-i} for fixed y^i, and has a finite upper bound.

(A3) π^i is supermodular in y^i, for fixed y^{-i}. For all strategy profiles y and y' in S,

$$\pi^i(y) + \pi^i(y') \le \pi^i(supremum\{y, y'\}) + \pi^i(infimum\{y, y'\}). \tag{A.11}$$

(A4) π^i has increasing differences in y^i and y^{-i}. For all $y^i \ge y^{i'}$, the difference $\pi^i(y^i, y^{-i}) - \pi^i(y^{i'}, y^{-i})$ is nondecreasing in y^{-i}.
In addition, it is sometimes useful to use a fifth assumption:

(A5) π^i has increasing differences in y^i and z for fixed y^{-i}; for all $y^i \ge y^{i'}$, the difference $\pi^i(y^i, y^{-i}, z) - \pi^i(y^{i'}, y^{-i}, z)$ is nondecreasing with respect to z.

The conditions for **smooth supermodularity** are

A1$'$ The strategy set is an interval in $R^{\bar{s}^i}$:

$$S^i = [\underline{y^i}, \overline{y^i}]. \tag{A.12}$$

A2$'$ π^i is twice continuously differentiable on S^i.

A3′   (Supermodularity) Increasing one component of player i's strategy does not decrease the net marginal benefit of any other component: for all i, and all s and t such that $1 \leq s < t \leq \overline{s}^i$,

$$\frac{\partial^2 \pi^i}{\partial y_s^i \partial y_t^i} \geq 0. \tag{A.13}$$

A4′   (Increasing differences in one's own and other strategies) Increasing one component of i's strategy does not decrease the net marginal benefit of increasing any component of player j's strategy: for all $i \neq j$, and all s and t such that $1 \leq s \leq \overline{s}^i$ and $1 \leq t \leq \overline{s}^j$,

$$\frac{\partial^2 \pi^i}{\partial y_s^i \partial y_t^j} \geq 0. \tag{A.14}$$

The fifth assumption becomes

A5′   (Increasing differences in parameters) Increasing parameter z does not decrease the net marginal benefit to player i of any component of his own strategy: for all i, and all s such that $1 \leq s \leq \overline{s}^i$,

$$\frac{\partial^2 \pi^i}{\partial y_s^i \partial z} \geq 0. \tag{A.15}$$

Theorem A.1   *If the game is supermodular, there exists a largest and smallest Nash equilibrium in pure strategies.*

Theorem A.1 is useful because it shows (1) existence of an equilibrium in pure strategies, and (2) if there are at least two equilibria (note that the largest and smallest equilibria might be the same strategy profile), then two of them can be ranked in the magnitudes of the components of each player's equilibrium strategy.

Theorem A.2   *If the game is supermodular and assumption* (A5′) *is satisfied, then the largest and smallest equilibria are nondecreasing functions of the parameter z.*

Theorem A.3   *If a game is supermodular, then for each player there is a largest and smallest serially undominated strategy, where both of these strategies are pure.*

Theorem A.4   *Let $\underline{y}^i$ denote the smallest element of player i's strategy set S^i in a supermodular game. Let y^* and $y^{*'}$ denote two equilibria, with $y^* \geq y^{*'}$, so y is the "big" equilibrium. Then,*

1   *If $\pi^i(\underline{y}^i, y^{-i})$ is increasing in y^{-i}, then $\pi^i(y^*) \geq \pi^i(y^{*'})$.*
2   *If $\pi^i(\underline{y}^i, y^{-i})$ is decreasing in y^{-i}, then $\pi^i(y^*) \leq \pi^i(y^{*'})$.*
3   *If the condition in* (1) *holds for a subset N_1 of players, and the condition in* (2) *holds for the remainder of the players, then the big equilibrium y^* is the best equilibrium for players in N_1 and the worst for the remaining player, and the small equilibrium $y^{*'}$ is the worst equilibrium for players in N_1 and the best for the remaining players.*

The theorems here are taken from Milgrom & Roberts (1990). Theorem A.1 is their corollary to theorem 5. Theorem A.2 is their theorem 6 and corollary. Theorem A.3 is their
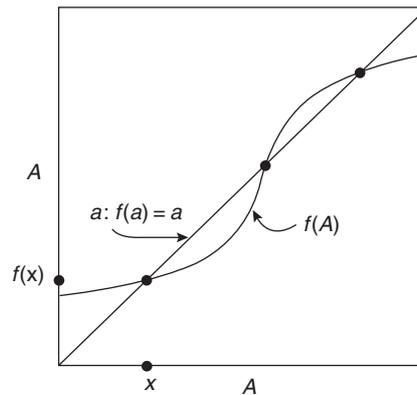
Figure A.3   A mapping with three fixed points.

theorem 5, and Theorem A.4 is their theorem 7. For more on supermodularity, see Milgrom & Roberts (1990), Fudenberg & Tirole (1991, pp. 489–97), or Vives's 2005 survey article. For a mathematician's view, see Topkis's 1998 *Supermodularity and Complementarity*.

A.7   Fixed Point Theorems

Fixed points theorems say that various kinds of mappings from one set to another result in at least one point being mapped back onto itself. The most famous fixed point theorem is Brouwer's Theorem, illustrated in figure A.3. I will use a formulation from page 952 of Mas-Colell, Whinston, & Green (1994).

The Brouwer Fixed Point Theorem

Suppose that set A in R^N is nonempty, compact, and convex; and that $f: A \rightarrow A$ is a continuous function from A into itself. ("Compact" means closed and bounded, in Euclidean space.) Then f has a fixed point; that is, there is an x in A such that $x = f(x)$.

The usefulness of fixed point theorems is that an equilibrium is a fixed point. Consider equilibrium prices. Let p be a point in N-space consisting of one price for each good. Let P be the set of all possible price points. P will be convex and compact if we limit it to finite prices. Agents in the economy look at p and make decisions about consumption and output. These decisions change p into $f(p)$. An equilibrium is a point p^* such that $f(p^*) = p^*$. If you can show that f is continuous, you can show that p^* exists.

This is also true for Nash equilibria. Let s be a point in *N-space* consisting of one strategy for each player – a strategy profile. Let S be the set of all possible strategy profiles. This will be compact and convex if we allow mixed strategies (for convexity) and if strategy sets are closed and bounded. Each strategy profile s will cause each player to react by choosing his best response $f(s)$. A Nash equilibrium is s^* such that $f(s^*) = s^*$. If you can show that f is continuous – which you can do if payoff functions are continuous – you can show that s^* exists.

The Brouwer theorem is useful in itself, and conveys the intuition of fixed point theorems, but to prove existence of prices in general equilibrium and existence of Nash equilibrium in game theory requires the Kakutani fixed point theorem. That is because the mappings involved are not one-to-one functions, but one-to-many-point correspondences. In general equilibrium, one firm might be indifferent between producing various amounts of output. In game theory, one player might have two best responses to another player's strategy.

The Kakutani Fixed Point Theorem

(Kakutani [1941])

Suppose that set A in R^N is a nonempty, compact, convex set and that $f: A \to A$ is an upper hemicontinuous correspondence from A into itself, with the property that the set $f(x)$ is nonempty and convex for every x. Then f has a fixed point; that is, there is an x in A such that x is one element of $f(x)$.

Other fixed point theorems exist for other kinds of mappings – for example, for a mapping from a set of functions back into itself. In deciding which theorem to use, care must be taken to identify the mathematical nature of the set of strategy profiles and the smoothness of the best response functions.

*A.8   Genericity

Suppose we have a space X consisting of the interval between 0 and 100 on the real line, $[0, 100]$, and a function f such that $f(x) = 3$ except that $f(15) = 5$. We can then say any of the following:

1 $f(x) = 3$ except on a set of measure zero.
2 $f(x) = 3$ except on a null set.
3 Generically, $f(x) = 3$.
4 $f(x) = 3$ almost always.
5 The set of x such that $f(x) = 3$ is dense in X.
6 The set of x such that $f(x) = 3$ has full measure.

These all convey the idea that if parameters are picked using a continuous random density, $f(x)$ will be 3 with probability one, and any other value of $f(x)$ is very special in that sense. If you start with point x, and add a small random perturbation ϵ, then with probability one, $f(x + \epsilon) = 3$. So unless there is some special reason for x to take the particular value of 15, you can count on observation $f(x) = 3$.

Statements like these always depend on the definition of the space X. If, instead, we took a space Y consisting of the integers between 0 and 100, which is $0, 1, 2, \ldots, 100$, then it is not true that "$f(x) = 3$ except on a set of measure zero." Instead, if x is chosen randomly, there is a $1/101$ probability that $x = 15$ and $f(x) = 5$.

The concept of "a set of measure zero" becomes more difficult to implement if the space X is not just a finite interval. I have not defined the concept in these notes; I have just pointed to usage. This, however, is enough to be useful to you. A course in real analysis would teach you the definitions. As with the concepts of "closed" and "bounded," complications can

arise even in economic applications because of infinite spaces and in dealing with spaces of functions, game tree branchings, or other such objects.

Now, let us apply the idea to games. Here is an example of a theorem that uses genericity.

Theorem A.5   "*Generically, all finite games of perfect information have a unique subgame perfect equilibrium.*"

Proof.   A game of perfect information has no simultaneous moves, and consists of a tree in which each player moves in sequence. Since the game is finite, each path through the tree leads to an end node. For each end node, consider the decision node just before it. The player making the decision there has a finite number N of choices, since this is a finite game. Denote the payoffs from these choices as $(P_1, P_2, \ldots, P_N)$. This set of payoffs has a unique maximum, because generically no two payoffs will be equal. (If they were, and you perturbed the payoffs a little, with probability one they would no longer be equal, so games with equal payoffs have measure zero.) The player will pick the action with the biggest payoff. Every subgame perfect equilibrium must specify that the players choose those actions, since they are the unique Nash strategies in the subgames at the end of the game.

Next, consider the next-to-last decision nodes. The player making the decision at such a node has a finite number of choices, and using the payoffs determined from the optimal choice of final moves, he will find that some move has the maximum payoff. The subgame perfect equilibrium must specify that move. Continue this procedure until you reach the very first move of the game. The player there will find that some one of his finite moves has the largest payoff, and he will pick that one move. Each player will have one best action choice at each node, and so the equilibrium will be unique. QED.

Genericity entered this as the condition that we are ignoring special games in the theorem's statement – games that have tied payoffs. Whether those are really special or not depends on the context.

*A.9   Discounting

A model in which the action takes place in real time must specify whether payments and receipts are valued less if they are made later, that is, whether they are **discounted**. Discounting is measured by the discount rate or the discount factor.

> The **discount rate**, r, is the extra fraction of a unit of value needed to compensate for delaying receipt by one period.

> The **discount factor**, δ, is the equivalent in present units of value of one unit to be received one period from the present.

The discount rate is analogous to the interest rate, and in some models the interest rate determines the discount rate. The discount factor represents exactly the same idea as the discount rate, and $\delta = 1/1 + r$. Models use r or δ depending on notational convenience. Not discounting is equivalent to $r = 0$ and $\delta = 1$, so the notation includes zero discounting as a special case.

Whether to put discounting into a model involves two questions. The first is whether the added complexity will be accompanied by a change in the results or by a surprising demonstration of no change in the results. A second, more specific question is whether the events of the model occur in real time, so that discounting is appropriate. The bargaining game of Alternating Offers from section 12.3 can be interpreted in two ways. One way is that the players make all their offers and counteroffers between dawn and dusk of a single day, so essentially no real time has passed. The other way is that each offer consumes a week of time, so that the delay before the bargain is reached is important to the players. Discounting is appropriate only in the second interpretation.

Discounting has two important sources: time preference and a probability that the game might end, represented by the rate of time preference, ρ, and the probability each period that the game ends, θ. It is usually assumed that ρ and θ are constant. If they both take the value zero, the player does not care whether his payments are scheduled now or ten years from now. Otherwise, a player is indifferent between $x/1 + \rho$ now and x guaranteed to be paid one period later. With probability $(1 - \theta)$ the game continues and the later payment is actually made, so the player is indifferent between $(1 - \theta)x/(1 + \rho)$ now and the promise of x to be paid one period later contingent upon the game still continuing. The discount factor is therefore

$$\delta = \frac{1}{1 + r} = \frac{(1 - \theta)}{(1 + \rho)}. \tag{A.16}$$

Table A.2 summarizes the implications of discounting for the value of payment streams of various kinds. We will not go into how these are derived, but they all stem from the basic fact that a dollar paid in the future is worth δ dollars now. Continuous time models usually refer to rates of payment rather than lump sums, so the discount factor is not so useful a concept, but discounting works the same way as in discrete time except that payments are continuously compounded. For a full explanation, see a finance text (e.g., Appendix A of Copeland & Weston [1988]).

The way to remember the formula for an annuity over a period of time is to use the formulas for a payment at a certain time in the future and for a perpetuity. A stream of x paid at the end of each year is worth x/r. A payment of Y at the end of period T has a present value of $-Y/(1 + r)^T$. Thus, if at the start of period T you must pay out a perpetuity of x at the end of each year, the present value of that payment is $(x/r)(1/1 + r)^T$. One may also view a stream of payments each year from the present until period T as the same thing as owning a perpetuity but having to give away a perpetuity in period T. This leaves a present value of $(x/r)(1 - (1/1 + r)^T)$, which is the second formula for an annuity given in table A.2. Figure A.4 illustrates this approach to annuities and shows how it can also be used to value a stream of income that starts at period S and ends at period T.
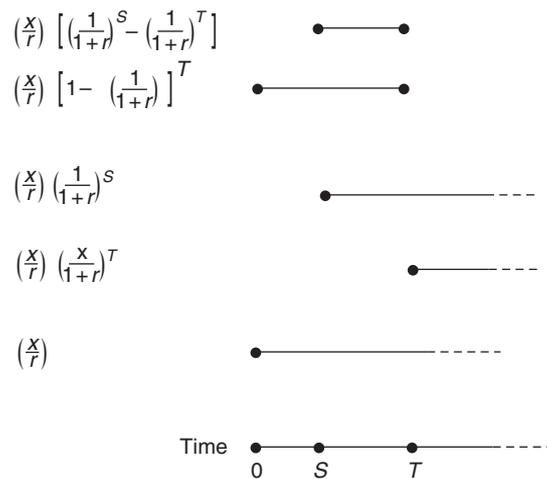
Discounting will be left out of most dynamic games in this book, but it is an especially important issue in infinitely repeated games, and is discussed further in section 5.2.

*A.10   Risk

We say that a player is **risk-averse** if his utility function is strictly concave in money, which means that he has diminishing marginal utility of money. He is **risk-neutral** if his utility

Table A.2   Discounting

Payoff Stream	Discounted Value	
	r-notation (discount rate)	δ-notation (discount factor)
x at the end of one period	$\dfrac{x}{1+r}$	δx
x at the end of each period in perpetuity	$\dfrac{x}{r}$	$\dfrac{\delta x}{1-\delta}$
x at the start of each period in perpetuity	$x + \dfrac{x}{r}$	$\dfrac{x}{1-\delta}$
x at the end of each period up through T (first formula)	$\displaystyle\sum_{t=1}^{T} \dfrac{x}{(1+r)^t}$	$\displaystyle\sum_{t=1}^{T} \delta^t x$
x at the end of each period up through T (second formula)	$\dfrac{x}{r}\left(1 - \dfrac{1}{(1+r)^T}\right)$	$\dfrac{\delta x}{1-\delta}(1-\delta^T)$
x at time t in continuous time	xe^{-rt}	—
Flow of x per period up to time T in continuous time	$\displaystyle\int_0^T xe^{-rt}\,dt$	—
Flow of x per period in perpetuity, in continuous time	$\dfrac{x}{r}$	—

$\left(\frac{x}{r}\right)\left[\left(\frac{1}{1+r}\right)^S - \left(\frac{1}{1+r}\right)^T\right]$ •————•

$\left(\frac{x}{r}\right)\left[1 - \left(\frac{1}{1+r}\right)\right]^T$ •—————•

$\left(\frac{x}{r}\right)\left(\frac{1}{1+r}\right)^S$ •—————- - -

$\left(\frac{x}{r}\right)\left(\frac{x}{1+r}\right)^T$ •———- - -

$\left(\frac{x}{r}\right)$ •————- - - - -

Time •—•————•——- - - -
0   S   T

Figure A.4   Discounting.

function is linear in money. The qualifier "in money" is used because utility may be a function of other variables too, such as effort.

We say that probability distribution F **dominates** distribution G in the sense of **first-order stochastic dominance** if the cumulative probability that the variable will take a value less than x is greater for G than for F, that is, if

$$\text{for any } x, \quad F(x) \le G(x), \tag{A.17}$$

and (A.17) is a strong inequality for at least one value of x. The distribution F dominates G in the sense of **second-order stochastic dominance** if the area under the cumulative distribution G up to $G(x)$ is greater than the area under F, that is, if

$$\text{for any } x, \quad \int_{-\infty}^{x} F(y)dy \le \int_{-\infty}^{x} G(y)dy, \tag{A.18}$$

and (A.18) is a strong inequality for some value of x. Equivalently, F dominates G if, limiting U to increasing functions for first-order dominance and increasing concave functions for second-order dominance,

$$\text{for all functions } U, \quad \int_{-\infty}^{+\infty} U(x)dF(x) > \int_{-\infty}^{+\infty} U(x)dG(x). \tag{A.19}$$

If F is a first-order dominant gamble, it is preferred by all players; if F is a second-order dominant gamble, it is preferred by all risk-averse players. If F is first-order dominant it is second-order dominant, but not vice versa.

Milgrom (1981b) has used stochastic dominance to carefully define what we mean by **good news**. Let θ be a parameter about which the news is received in the form of message x or y, and let utility be increasing in θ. The message x is more favorable than y (is "good news") if for every possible nondegenerate prior for $F(\theta)$, the posterior $F(\theta|x)$ first-order dominates $F(\theta|y)$.

Rothschild & Stiglitz (1970) shows how two gambles can be related in other ways equivalent to second-order dominance, the most important of which is the **mean-preserving spread**. Informally, a mean-preserving spread is a density function which transfers probability mass from the middle of a distribution to its tails. More formally, for discrete distributions placing sufficient probability on the four points a_1, a_2, a_3, and a_4,

*A **mean-preserving spread** is a set of four locations $a_1 < a_2 < a_3 < a_4$ and four probabilities $\gamma_1 \ge 0, \gamma_2 \le 0, \gamma_3 \le 0, \gamma_4 \ge 0$ such that $-\gamma_1 = \gamma_2$, $\gamma_3 = -\gamma_4$, and $\sum_i \gamma_i a_i = 0$.*

Figure A.5 shows how this works. Panel (a) shows the original distribution with solid bars. The mean is $0.1(2) + 0.3(3) + 0.3(4) + 0.2(5) + 0.1(6)$, which is 3.9. The spread has mean $0.1(1) - 0.3(.3) - 0.1(4) + 0.2(6)$, which is 0, so it is mean-preserving. Panel (b) shows the resulting spread-out distribution.

The definition can be extended to continuous distributions, and can be alternatively defined by taking probability mass from one point in the middle and moving it to the sides; panel (c) of figure A.5 shows an example. See Rasmusen & Petrakis (1992), which also, with Leshno, Levy, & Spector (1997), fixes an error in the original Rothschild & Stiglitz proof.
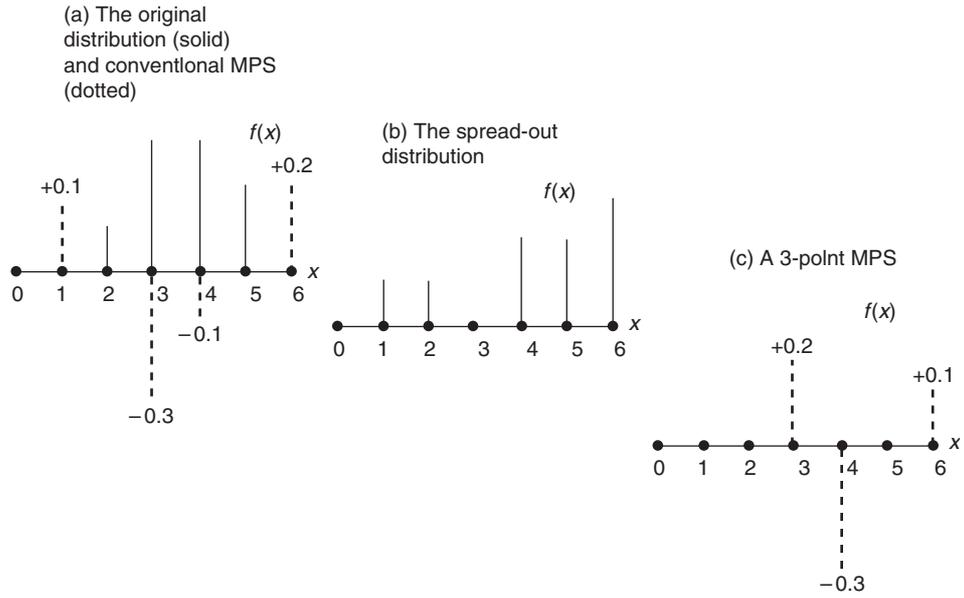
**Figure A.5**   Mean-preserving spreads.

Hazard Rates

The hazard rate has been important in the analysis above. Suppose we have a density $f(v)$ for a buyer's value for an object being sold, with cumulative distribution $F(v)$ on support $[\underline{v}, \overline{v}]$. The hazard rate $h(v)$ is defined as

$$h(v) = \frac{f(v)}{1 - F(v)} \tag{A.20}$$

What this means is that $h(v)$ is the probability density of v for the distribution which is like $F(v)$ except cutting off the lower values, so its support is limited to $[v, \overline{v}]$. In economic terms, $h(v)$ is the probability density for the valuing equalling v given that we know that value equals at least v.

For most distributions we use, the hazard rate is increasing, including the uniform, normal, logistic, and exponential distributions, and any distribution with increasing density over its support (see Bagnoli & Bergstrom [1994]). Figure A.6 shows three of them.
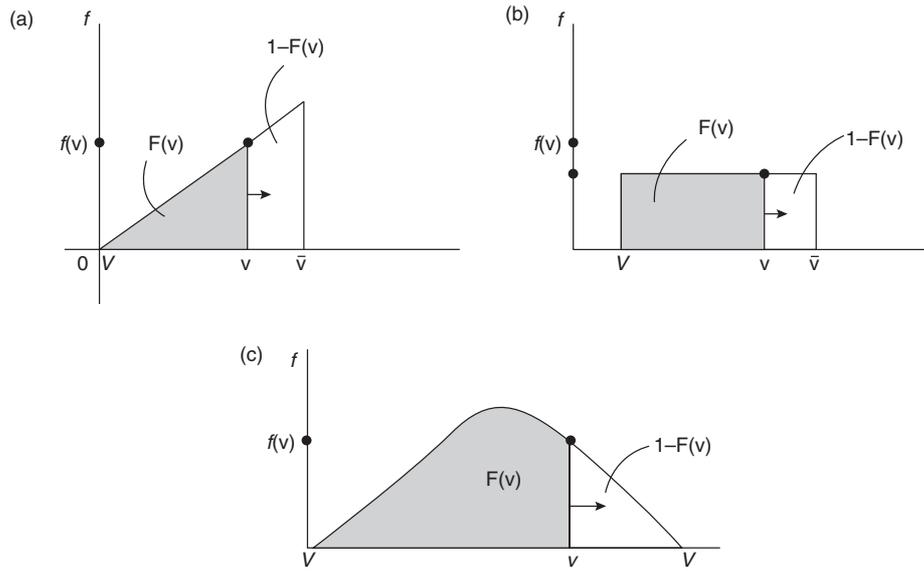
Figure A.6   Three densities to illustrate increasing hazard rates $f(v)/(1 - f(v))$.