Math

Nathan Ueda

January 2024

1 Stochastic Matrices

1.1 Stochastic Matrices

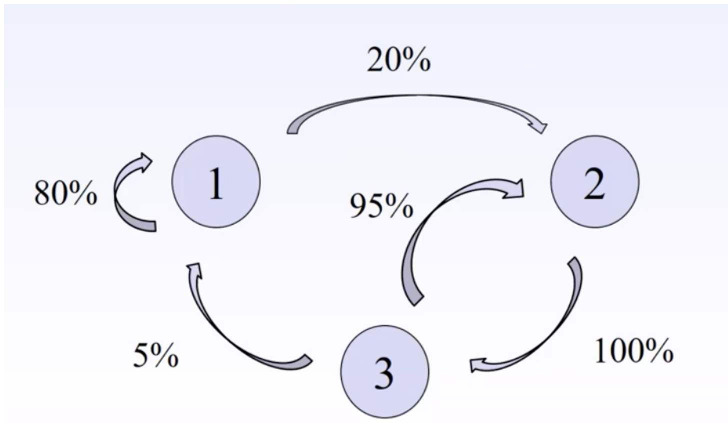
- Informal definition: A square matrix that represent transitions between different random states. Elements within a stochastic matrix represent probabilities. This implies that each element must be nonnegative and each row must sum to 1.
- Formal definition: A matrix $\Phi \in R^{N \times N}$ is a stochastic matrix if:
  - all elements are positive: $\Phi_{ij} \geq 0$ for all i, j
  - each row sums to 1: $\Sigma_j \Phi_{ij} = 1$ for all i

1.1.1 Intepretations

- $P(s_{t+1} = n | s_t = m) = \Phi_{mn}$: This says that the element Φ_{mn} represents the probability of the next state, s_{t+1} being equal to n given that the current state, s_t is the mth state for all possible states $s \in \{1, \ldots, N\}$.

1.1.2 Examples

1. 3-State Markov Process: Stochastic Matrix $\Phi = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0 & 1 \\ 0.05 & 0.95 & 0 \end{bmatrix}$



Each node is a state. Each edge represents the probability of transitioning from some state to another.

1.2 Discrete Kolmogorov Backward Equations

- Informal definition: A formula to represent the expected value of some payoff function (i.e. for an option) depending upon what state we are in the future.

- Formal definition: For function $F : \{1, \dots, N\} \to R$, define $F(j) = f_j, f \in \mathbb{R}^N$.

  - F is the 'payoff function' and f_j is the payoff, given we are in state j at some future specified time.

1.2.1 Intepretations

1. $\Phi^2 = \Phi\Phi$: The transition probabilities between two periods, t and $t+2$.

2. Φ^n: The transition probabilities between n periods, t and $t+n$.

3. $[\Phi^{T-t}]_{ij} = P(s_T = j | s_t = i)$: The probability of being in state j after t periods, given that we are in state i currently (at time t) is the ij-th element of the stochastic matrix Φ^{T-t}.

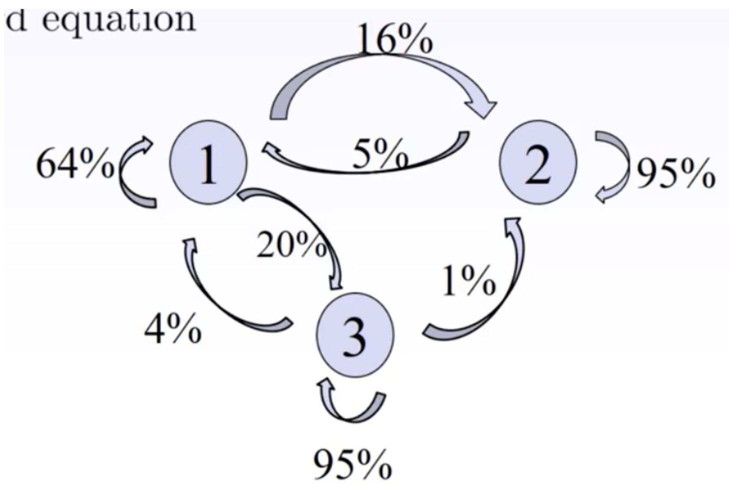
4. $E[F(s_T)|s_t = j] = [\Phi^{T-t} f]_j$: The expected payoff at period T, given that we are currently in state $s_t = j$ at period t, is the j-th element of the vector resulting from taking the product of Φ^{T-t} and the payoff function f.

   - $[\Phi^{T-t} f]_j = \Phi^{T-t}_{j1} f_1 + \cdots + \Phi^{T-t}_{jN} f_N$

1.2.2 Examples

1. 3-State Markov Process: Stochastic Matrix $\Phi = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0 & 1 \\ 0.05 & 0.95 & 0 \end{bmatrix}$,

   $\Phi^2 = \begin{bmatrix} 0.64 & 0.16 & 0.2 \\ 0.05 & 0.95 & 0 \\ 0.04 & 0.01 & 0.95 \end{bmatrix}$



Each node is a state. Each edge represents the probability of ending up in the state at the end of the arrow, given we are in the state at the tail of the arrow 2 periods prior.

1.3 Discrete Kolmogorov Forward Equations

- Informal definition: An equation useful in determining the distribution of us being in a specific state t periods into the future.
- Formal definition:
  - Probability vector: a vector $p \in \mathbb{R}^N$ such that:
    * $p_i \geq 0$
    * $\Sigma_i p_i = 1$
  - Stationary distribution: a probability vector for a stochastic matrix, Φ, such that $p = \Phi^* p$ (i.e. p is a left eigenvector of Φ^* with an eigenvalue of 1) (Φ^* is used to denote the transpose).
  - Discrete Kolmogorov Forward Equation: if the probability vector at t is p^t, then at time T, the probability vector for being in any given state is $(\Phi^*)^{T-t} p^t$.

- To ensure a stochastic matrix has a unique stationary distribution, two criteria must be fulfilled:

  1. The matrix must be **irreducible**. As stochastic matrix is said to be irreducible if, for each i, j, there exists $k > 0$ such that $(\Phi^k)_{ij} > 0$. In other words, there exists some path that has a non-zero probability between any two states.
  2. The Markov chain is **aperiodic**. The **period** of state i, is k_i, where k_i is the greatest common divisor of the set $\{n|(\Phi^n)_{ii} > 0\}$. If $k_i = 1$ for $i = 1, \ldots, n$, then Φ is aperiodic. In other words, to be aperiodic means there is no cyclicality in the number of steps it takes to get back to some starting state.
  3. **Perron-Frobenius Theorem**: Consider an irreducible, aperiodic stochastic matrix Φ. Then Φ has a unique stationary distribution with strictly positive elements.
     - All other left eigvenvalues of $\Phi, \lambda_2, \ldots, \lambda_N$, have $|\lambda_i| < 1$.
     - All other left eigenvectors of Φ have at least one nonpositive element.

1.3.1 Intepretations

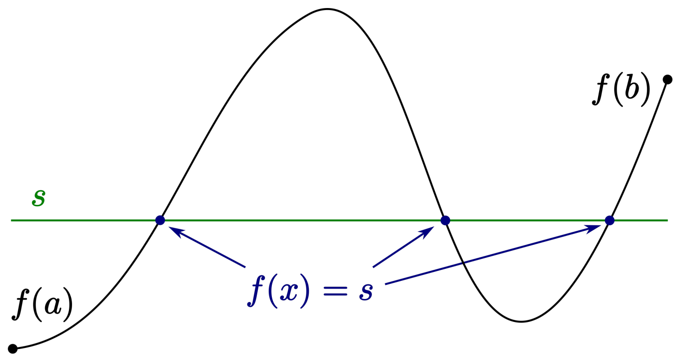
1. $\Phi^{1000} = \begin{bmatrix} 0.111 & 0.444 & 0.444 \\ 0.111 & 0.444 & 0.444 \\ 0.111 & 0.444 & 0.444 \end{bmatrix}$: This is the transition matrix 1000 periods into the future. Each column has the same values. This means regardless of which state we start in, we have the same probability of ending up in any state 1000 periods later. This implies there is a limiting distribution $\mathbf{p} = (0.111, 0.444, 0.444)'$ regardless of the initial state. Note: there is not always a limiting distribution.

1.4 Multidimensional Calculus

1.4.1 Intermediate Value Therorem

1. Informal Definition: Given some continuous function that maps a point $a \to f(a)$ and $b \to f(b)$, then there exists some point s where $a \leq s \leq b$, such that $f(a) \leq f(s) \leq f(b)$.

2. Formal Definition: Consider a continuous function $f : [a, b] \to R$ and a number $x \in [f(a), f(b)]$. Then $\exists\ s \in [a, b]$ such that $f(s) = x$.



1.4.2 Contraction Mapping Theorem

- Informal Definition: A theorem that, if it holds, guarantees the existence and uniqueness of a fixed point for a complete metric space.
- Formal Definition: If X is a complete metric space and $f : X \to X$ is a contraction, then f has a unique fixed point $x \in X$.
  - A function $f : X \to X$ is called a contraction if $d(f(x), f(y)) \leq k \times d(x, y)\ \forall\ x, y \in X$, for some $k < 1$. In other words, a function is a contraction if it decreases distances between every pair of points.

1.4.3 Partial Derivatives

- For a function $f : \mathbb{R}^N \to R^M$, define the partial derivative: $a_{nm} = \frac{\partial f(x)_m}{\partial x_n} = D_n(f)_m$.
- Theorem: If f is differentiable at a point x, then the total derivative is given by $A \in R^{M\times N}$, where $A = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{bmatrix}$
- Caveat: Partial derivatives may exist even when the total derivative does not.
- When $M = 1$, we have a gradient operator: $\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix}^T$
- The Hessian matrix is a square matrix of second order partial derivatives:

$$H = \nabla^2 = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

1.5 Optimization

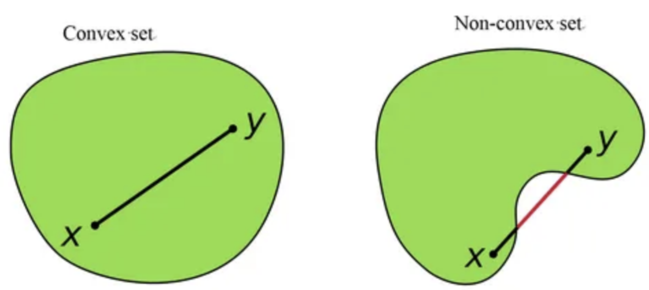
1.5.1 Univariate General Existence Result Theorem

- Informal Definition: If we can make a few assumptions about a function, namely that it has a well-defined domain, is closed, and is bounded, then the existence of a maximum and a minimum to a function is guaranteed.
- Formal Definition: Suppose $f : E \to R$ is a continuous function from a closed, bounded set $E \subset \mathbb{R}^N$. Define:
  - $M = \sup_{x \in E} f(x)$
  - $m = \inf_{x \in E} f(x)$

  Then there exists points $\underline{x} \in E, \bar{x} \in E$ such that $f(\underline{x}) = m$ and $f(\bar{x} = M)$ (theorem 4.16 in Rudin).

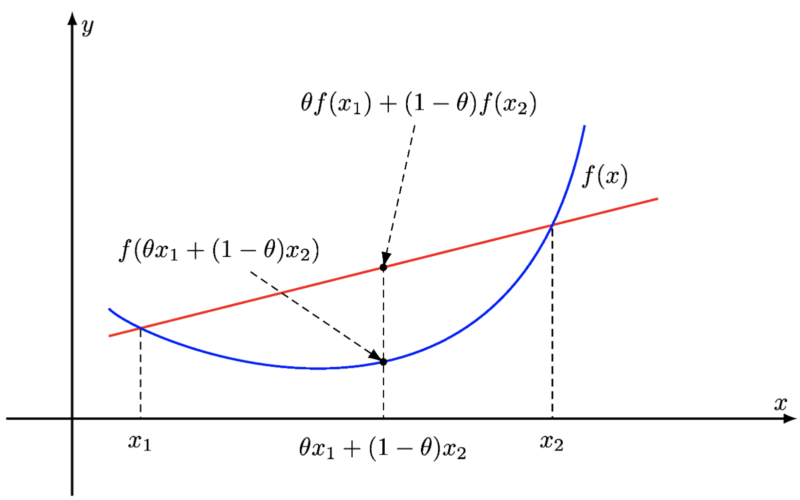
1.5.2 Convex Set

- Informal Definition: A set is **convex** if for every $\lambda \in (0, 1)$ and any two points x, y in the set, the weighted average that puts λ weight on x and $1 - \lambda$ weight on y is also in the set.
- Formal Definition: A set $E \subset \mathbb{R}^N$ is said to be **convex** if $\forall \lambda \in (0, 1), \forall x \in E, \forall y \in E : \lambda x + (1 - \lambda)y \in E$

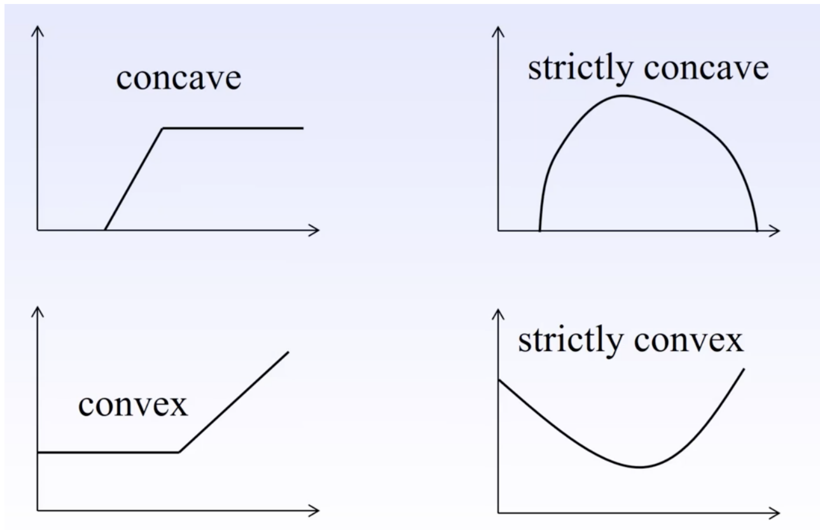


1.5.3 Concave and Convex Functions

- Informal Definition: A function on a convex set is
  - **concave** if for every $\lambda \in (0,1)$ and any two points x, y in the set, the function evaluated at $\lambda x + (1-\lambda)y$ (a point between x and y) is greater than or equal to the weighted average of the values at those points.
  - **convex** if for every $\lambda \in (0,1)$ and any two points x, y in the set, the function evaluated at $\lambda x + (1-\lambda)y$ (a point between x and y) is less than or equal to the weighted average of the values at those points.
  - **stricly concave** if for every $\lambda \in (0,1)$ and any two points x, y in the set, the function evaluated at $\lambda x + (1-\lambda)y$ (a point between x and y) is greater than to the weighted average of the values at those points.
  - **strictly convex** if for every $\lambda \in (0,1)$ and any two points x, y in the set, the function evaluated at $\lambda x + (1-\lambda)y$ (a point between x and y) is less than to the weighted average of the values at those points.
- Formal Definition: A function $f : E \to \mathbb{R}^N$, where $E \subset \mathbb{R}^N$ is convex, is said to be:
  - **concave** if $\forall \lambda \in (0,1), x, y \in E, f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y)$
  - **convex** if $\forall \lambda \in (0,1), x, y \in E, f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
  - **stricly concave**: replace $\geq$ with $>$
  - **stricly convex**: replace $\leq$ with $<$



Graph of a convex function. The line segment between any two points on the graph lies above the graph.



1.6 Smooth Functions

1.6.1 Smooth Functions in One Dimension

- Informal Definition: For a 1-D function from an interval to the real line, if that function is twice continuously differentiable, then the function is:
  - concave if and only if the second derivative is less than or equal to 0.
  - concave if and only if the second derivative is less than 0.
  - convex if and only if the second derivative is greater than or equal to 0.
  - convex if and only if the second derivative is greater than 0.
- Formal Definition: If $f : [a, b] \to \mathbb{R}$ is C^2, then f is
  - concave $\iff$ $f'' \leq 0$
  - strictly concave $\iff$ $f'' < 0$
  - convex $\iff$ $f'' \geq 0$
  - strictly convex $\iff$ $f'' > 0$

1.6.2 Conditions for Optimality in One Dimension Theorem

- Informal Definition: If we have a function that is once continuously differentiable, then for some value x^* in the inverval $[a, b]$ such that $f'(x^*) = 0$ and f is:
  - concave, then x^* is a maxmium

- strictly concave, then x^* is the unique maxmium
  - convex, then x^* is a minimum
  - strictly convex, then x^* is the unique minimum

- Formal Definition: Consider a C^1 function $f : [a, b] \to \mathbb{R}$, which satisfies $f'(x^*) = 0$ for some $x^* \in [a, b]$. If f is:
  - concave, then x^* is a maxmium
  - strictly concave, then x^* is the unique maxmium
  - convex, then x^* is a minimum
  - strictly convex, then x^* is the unique minimum

1.6.3 Smooth Functions in Multiple Dimensions

- Formal Definition: If $f : E \to \mathbb{R}$ is C^2
  - f is concave $\iff H_f$ is negative semidefinite
  - f is strictly concave $\iff H_f$ is negative definite
  - f is convex $\iff H_f$ is positive semidefinite
  - f is strictly convex $\iff H_f$ is positive definite

1.6.4 Conditions for Optimality in Multiple Dimension Theorem

- Informal Definition: If we have a function that is twice continuously differentiable, then for some convex set $E \subset \mathbb{R}^N$ where the first order condition is satisfied (the gradient at some point x^* is equal to the 0 vector) for some $x^* \in E$, the function is
  - concave, then x^* is a maxmium
  - strictly concave, then x^* is the unique maxmium
  - convex, then x^* is a minimum
  - strictly convex, then x^* is the unique minimum

- Formal Definition: Consider a C^2 function $f : E \to \mathbb{R}$ for a convex set $E \subset \mathbb{R}^N$, where f satisfies $\nabla f(x^*) = 0$ for some $x^* \in E$. If f is:
  - concave, then x^* is a maxmium
  - strictly concave, then x^* is the unique maxmium
  - convex, then x^* is a minimum
  - strictly convex, then x^* is the unique minimum

1.6.5 Lagrange Optimality Condition

- Informal Definition:

- Formal Definition