An Introduction to Computational Macroeconomics

Dynamic Programming: Chapter 2

John Stachurski

June – July 2022

Markov Dynamics

Our next task is to review Markov dynamics

- An **essential workhorse** for countless models in
  - economics
  - finance
  - operations research, etc.
- Very general – can handle most processes of interest
- Elegant theory
- Natural fit for dynamic programming (Markov decisions)

Topics

1. Nonnegative matrices
2. The Perron–Frobenius theorem
3. A lake model of employment flows
4. Markov chains
5. Stationarity and ergodicity
6. Approximation
7. Expectations

Reminders

Def. $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ is an **eigenpair** of $n \times n$ matrix A if

$$v \neq 0 \quad \text{and} \quad Av = \lambda v$$

The **eigenspace** of eigenvalue λ is

$$E_\lambda := \{w \in \mathbb{C}^n : w = 0 \text{ or } (\lambda, w) \text{ is an eigenpair of } A\}$$

Ex. Show that E_λ is a linear subspace of $\mathbb{C}^n$

Proof: If $v, w \in E_\lambda$ and $\alpha, \beta \in \mathbb{C}$, then

$$A(\alpha w + \beta v) = \alpha Aw + \beta Av = \alpha\lambda w + \beta\lambda v = \lambda(\alpha w + \beta v)$$

Reminders

Def. $(\lambda, v) \in \mathbb{C} \times \mathbb{C}^n$ is an **eigenpair** of $n \times n$ matrix A if

$$v \neq 0 \quad \text{and} \quad Av = \lambda v$$

The **eigenspace** of eigenvalue λ is

$$E_\lambda := \{w \in \mathbb{C}^n : w = 0 \text{ or } (\lambda, w) \text{ is an eigenpair of } A\}$$

Ex. Show that E_λ is a linear subspace of $\mathbb{C}^n$

Proof: If $v, w \in E_\lambda$ and $\alpha, \beta \in \mathbb{C}$, then

$$A(\alpha w + \beta v) = \alpha Aw + \beta Av = \alpha\lambda w + \beta\lambda v = \lambda(\alpha w + \beta v)$$

Implication: exists a continuum of eigenvectors paired with λ

So what can we say about uniqueness?

Let (λ, v) be an eigenpair for A

Def. v has (geometric) **multiplicity one** if $\dim E_\lambda = 1$

In other words,

$$w \in E_\lambda \implies w = \alpha v \text{ for some } \alpha \in \mathbb{C}$$

In a sense, there is "just one" eigenvector corresponding to λ, since any other is a scalar multiple

Nonnegative Matrices

Def. Matrix A is called

- **nonnegative**, and we write $A \geqslant 0$, if all elements of A are nonnegative
- **positive**, and we write $A \gg 0$, if every element of A is strictly positive
- **irreducible** if it is square, nonnegative and

$$\sum_{k \in \mathbb{N}} A^k \gg 0$$

Note: positive $\implies$ irreducible $\implies$ nonnegative

Let A be $n \times n$

It is not always true that $r(A)$ is an eigenvalue of A

Example. Let

$$A := \begin{pmatrix} -1 & 0 \\ 0 & 1/2 \end{pmatrix}$$

Set of eigenvalues (the **spectrum**) of A is $\sigma(A) = \{-1, 1/2\}$

Hence $r(A) = |-1| = 1 \notin \sigma(A)$

However, when $A \geqslant 0$, we have the following result

Theorem. (Perron–Frobenius) If $A \geqslant 0$, then $r(A)$ is an eigenvalue of A with nonnegative, real-valued right and left eigenvectors

In particular, there exists

- a nonnegative, nonzero column vector e s.t. $Ae = r(A)e$
- a nonnegative, nonzero row vector ε s.t. $\varepsilon A = r(A)\varepsilon$

If A **is irreducible**, then these eigenvalues are everywhere positive and have multiplicity of one

If A **is positive**, then with e and ε such that $\langle \varepsilon, e \rangle = 1$, we have

$$r(A)^{-t} A^t \to e\,\varepsilon \qquad (t \to \infty)$$

In this setting,

- $r(A)$ is also called the **dominant eigenvalue**
- e is called the **dominant right eigenvector**
- ε is called the **dominant left eigenvector**

Note also

$$\varepsilon A = r(A)\varepsilon \quad \iff \quad A^\top \varepsilon^\top = r(A)\varepsilon^\top$$

Hence $\varepsilon^\top$ is the dominant right eigenvector of $A^\top$

Since the dominant eigenvectors are only defined up to constant multiples, we often normalize so that $\langle \varepsilon, e \rangle = 1$

Let's check these results for arbitrary positive A

```
julia> using LinearAlgebra

julia> A = [0.3 0.9;
            1.0 0.1];

julia> λ_1, λ_2 = eigvals(A)
2-element Vector{Float64}:
 -0.7539392014169456
  1.1539392014169458

julia> rA = λ_2    # r(A) is the positive eigenvalue
1.1539392014169458
```

```julia
julia> right_evecs = eigvecs(A)
2×2 Matrix{Float64}:
 -0.649386  0.725426
  0.760459  0.6883

julia> e = right_evecs[:, 2]   # dominant right eigenvector
2-element Vector{Float64}:
 0.7254262498099013
 0.6882999027217298

julia> left_evs = eigvecs(A')  # transpose to get left eigenvector
2×2 Matrix{Float64}:
 -0.6883    0.760459
  0.725426  0.649386

julia> ϵ = left_evs[:, 2]'     # dominant left eigenvector
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.760459  0.649386
```

```
#  Checking the eigenpair relations

 julia> A * e
2-element Vector{Float64}:
 0.8370977873925273
 0.7942562400820743

julia> rA * e
2-element Vector{Float64}:
 0.8370977873925274
 0.7942562400820744


julia> ϵ * A
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.877524  0.749352

julia> rA * ϵ
1×2 adjoint(::Vector{Float64}) with eltype Float64:
 0.877524  0.749352
```

```
#  The matrix A is everywhere positive
#
#  Hence we expect, for large k,
#
#       r(A)^(-k) * A^k ≈ e ϵ


julia> k = 1000
1000

julia> rA^(-k) * A^k
2×2 Matrix{Float64}:
 0.552414  0.471728
 0.524142  0.447586

julia> e * ϵ
2×2 Matrix{Float64}:
 0.551657  0.471082
 0.523424  0.446972
```

Corollary: bounds on the spectral radius

Fix $n \times n$ matrix A and set

- $\mathrm{rs}_i(A) :=$ the i-th row sum of A and
- $\mathrm{cs}_j(A) :=$ the j-th column sum of A

Corollary. If $A \geqslant 0$, then

1. $\min_i \mathrm{rs}_i(A) \leqslant r(A) \leqslant \max_i \mathrm{rs}_i(A)$ and
2. $\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$

Ex. Prove this via the PF theorem

Proof for the column sum case

Fix $A \geqslant 0$ and let e be the dominant right eigenvector

We normalize e by setting $\mathbb{1}^\top e = \sum_j e_j = 1$

From $r(A)e = Ae$ we have

$$r(A) = r(A)\mathbb{1}^\top e = \mathbb{1}^\top (r(A)e) = \mathbb{1}^\top Ae = \sum_j \mathrm{cs}_j(A) e_j$$

Therefore, $r(A)$ is a weighted average of the column sums

Hence $\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$

Stochastic Matrices

Let P be a square matrix

Def. P is called **stochastic** if $P \geqslant 0$ and $P\mathbb{1} = \mathbb{1}$

Ex. Show that P is stochastic $\implies r(P) = 1$

Row vector ψ is called a **stationary distribution** of P if

$$\psi \geqslant 0, \quad \psi\mathbb{1} = 1 \quad \text{and} \quad \psi P = \psi$$

Stationary distributions very important for Markov dynamics...

Existence of Stationary Distributions

Let P be a stochastic matrix

Ex. Prove: P has at least one stationary distribution

Proof: By the PF theorem,

$\exists$ a nonzero, nonnegative row vector φ satisfying $\varphi P = \varphi$

Since φ is nonzero, $\varphi \mathbb{1} > 0$

Setting $\psi := \varphi / (\varphi \mathbb{1})$ gives the desired vector

Existence of Stationary Distributions

Let P be a stochastic matrix

Ex. Prove: P has at least one stationary distribution

Proof: By the PF theorem,

$$\exists \text{ a nonzero, nonnegative row vector } \varphi \text{ satisfying } \varphi P = \varphi$$

Since φ is nonzero, $\varphi \mathbb{1} > 0$

Setting $\psi := \varphi / (\varphi \mathbb{1})$ gives the desired vector

Uniqueness of Stationary Distributions

Ex. Prove: If P is also **irreducible**, then the stationary vector ψ is everwhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

Uniqueness of Stationary Distributions

Ex. Prove: If P is also **irreducible**, then the stationary vector ψ is everwhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

Uniqueness of Stationary Distributions

Ex. Prove: If P is also **irreducible**, then the stationary vector ψ is everwhere positive and unique

Proof of Positivity: See Perron–Frobenius theorem

Proof of Uniqueness: Let $\varphi \geqslant 0$ satisfy $\varphi \mathbb{1} = 1$ and $\varphi P = \varphi$

By the Perron–Frobenius theorem, $\varphi = \alpha \psi$ for some $\alpha > 0$

But then $1 = \varphi \mathbb{1} = \alpha \psi \mathbb{1} = \alpha$

Hence $\varphi = \psi$

```
julia> P = [0.2 0.8;
            0.1 0.9]
2×2 Matrix{Float64}:
    0.2  0.8
    0.1  0.9

julia> using QuantEcon

julia> mc = MarkovChain(P)
    Discrete Markov Chain
    stochastic matrix of type Matrix{Float64}:
    [0.2 0.8; 0.1 0.9]

julia> is_irreducible(mc)
true

julia> stationary_distributions(mc)
    1-element Vector{Vector{Float64}}:
    [0.1111111111111111, 0.8888888888888888]
```

Lake Model of Employment

An illustration of the Perron–Frobenius theorem

We analyze a model of employment and unemployment flows in a large population

The model is sometimes called a "lake model"

Two "pools" of workers:

- those who are currently employed and
- those who are currently unemployed but still seeking work

FP theorem helps us analyze dynamics

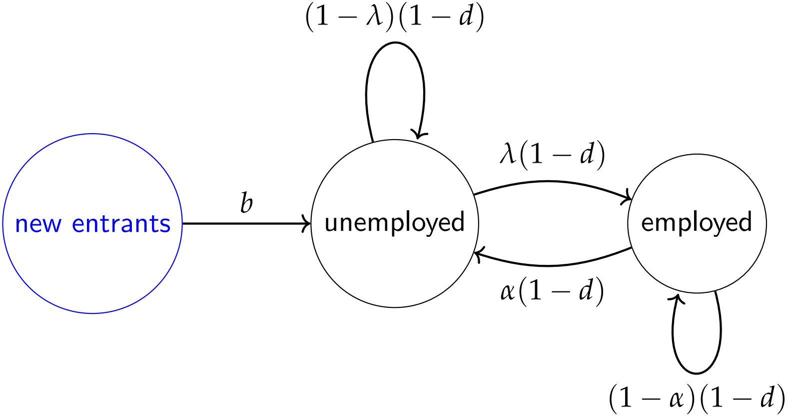
Workers

- **exit** the workforce at rate d
- **enter** the workforce at rate b
- **separate** from their jobs at rate α
- **find jobs** at rate λ

Assumptions:

- All parameters lie in $(0, 1)$
- New workers are initially unemployed

Transition rates:



Let

- $u_t :=$ number of **unemployed workers** at time t
- $e_t :=$ number of **employed workers**
- $n_t := e_t + u_t :=$ total **population** of workers

Dynamics are

$$u_{t+1} = (1-d)\alpha e_t + (1-d)(1-\lambda)u_t + bn_t$$

$$e_{t+1} = (1-d)(1-\alpha)e_t + (1-d)\lambda u_t$$

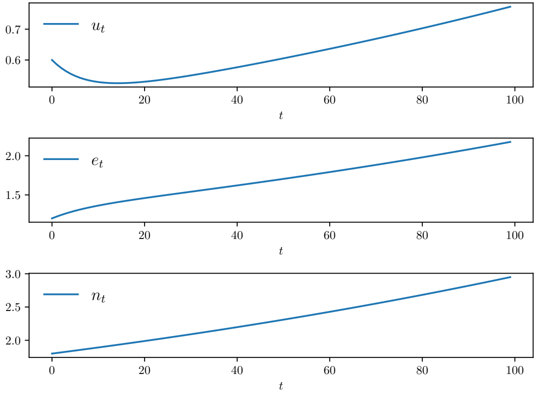


Figure: Example simulation when $b > d$ (population growth)

Can we say more about the dynamics of this system?

For example,

- what long run unemployment rate should we expect?
- do outcomes depend on the initial conditions u_0 and e_0?
- Or are there general statements we can make?

We define

$$x_t := \begin{pmatrix} u_t \\ e_t \end{pmatrix}$$

and

$$A := \begin{pmatrix} (1-d)(1-\lambda)+b & (1-d)\alpha+b \\ (1-d)\lambda & (1-d)(1-\alpha) \end{pmatrix}$$

Dynamics can now be written

$$x_{t+1} = Ax_t$$

Hence

$$x_t = A^t x_0 \quad \text{where} \quad x_0 = \begin{pmatrix} u_0 \\ e_0 \end{pmatrix}$$

Ex. With $g := b - d$, show that $n_{t+1} = (1+g)n_t$ for all t

Proof: The column sums of A are

$$(1-d)(1-\lambda) + b + (1-d)\lambda = 1 + g$$

and

$$(1-d)\alpha + b + (1-d)(1-\alpha) = 1 + g$$

From $x_{t+1} = Ax_t$ and $n_t = u_t + e_t$ we have

$$n_{t+1} = \mathbb{1}^\top x_{t+1} = \mathbb{1}^\top A x_t = (1+g)\mathbb{1}^\top x_t = (1+g)n_t$$

Ex. With $g := b - d$, show that $n_{t+1} = (1+g)n_t$ for all t

Proof: The column sums of A are

$$(1-d)(1-\lambda) + b + (1-d)\lambda = 1 + g$$

and

$$(1-d)\alpha + b + (1-d)(1-\alpha) = 1 + g$$

From $x_{t+1} = Ax_t$ and $n_t = u_t + e_t$ we have

$$n_{t+1} = \mathbb{1}^\top x_{t+1} = \mathbb{1}^\top A x_t = (1+g)\mathbb{1}^\top x_t = (1+g)n_t$$

Ex. Prove that $r(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant r(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = r(A)\mathbb{1}^\top$$

Ex. Prove that $r(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant r(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = r(A)\mathbb{1}^\top$$

Ex. Prove that $r(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant r(A) \leqslant 1 + g$

$$\text{PF theorem} \implies 1 + g \text{ is the dominant eigenvalue of } A$$

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = (1 + g \quad 1 + g) = r(A)\mathbb{1}^\top$$

Ex. Prove that $r(A) = 1 + g$

Proof: We know that

$$\min_j \mathrm{cs}_j(A) \leqslant r(A) \leqslant \max_j \mathrm{cs}_j(A)$$

Hence $1 + g \leqslant r(A) \leqslant 1 + g$

PF theorem $\implies$ $1 + g$ is the dominant eigenvalue of A

Ex. Show that $\mathbb{1}^\top := (1\ 1)$ is the dominant left eigenvector of A

Proof:

$$\mathbb{1}^\top A = \begin{pmatrix} 1+g & 1+g \end{pmatrix} = r(A)\mathbb{1}^\top$$

Ex. Prove that

$$\bar{x} := \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix}$$

with

$$\bar{u} := \frac{1+g-(1-d)(1-\alpha)}{1+g-(1-d)(1-\alpha)+(1-d)\lambda} \quad \text{and} \quad \bar{e} := 1-\bar{u}$$

is the dominant right eigenvector of A

Proof: Just show $A\bar{x} = (1+g)\bar{x}$

Ex. Prove that

$$\bar{x} := \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix}$$

with

$$\bar{u} := \frac{1 + g - (1-d)(1-\alpha)}{1 + g - (1-d)(1-\alpha) + (1-d)\lambda} \quad \text{and} \quad \bar{e} := 1 - \bar{u}$$

is the dominant right eigenvector of A

Proof: Just show $A\bar{x} = (1+g)\bar{x}$

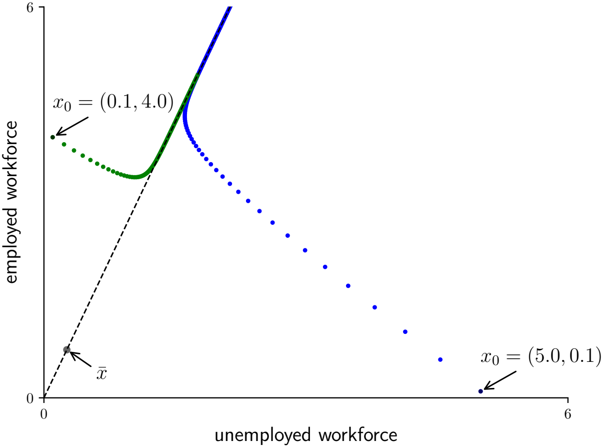
```julia
using LinearAlgebra

α, λ, d, b = 0.01, 0.1, 0.02, 0.025
g = b - d
A = [(1 - d) * (1 - λ) + b     (1 - d) * α + b;
     (1 - d) * λ               (1 - d) * (1 - α)]

ū = (1 + g - (1 - d) * (1 - α)) /
        (1 + g - (1 - d) * (1 - α) + (1 - d) * λ)

ē = 1 - ū
x̄ = [ū; ē]

println(isapprox(A * x̄, (1 + g) * x̄))   # prints true
```


6
0
0
6
employed workforce
unemployed workforce
$x_0 = (0.1, 4.0)$
$x_0 = (5.0, 0.1)$
$\bar{x}$

Let

$$D := \{x \in \mathbb{R}^2 : x = \alpha \bar{x} \text{ for some } \alpha > 0\}$$

- Shown as a dashed black line in the last figure
- The two time paths are of the form $(x_t)_{t \geqslant 0} = (A^t x_0)_{t \geqslant 0}$
- In both cases, the paths converge to D over time

Suggests all paths are "eventually almost" multiples of $\bar{x}$

How can we explain this strong regularity?

From the Perron–Frobenius theorem, since $A \gg 0$, we have

$$A^t \approx r(A)^t \cdot \bar{x}\mathbb{1}^\top = (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \quad \text{for large } t$$

Hence, $\forall\, x_0 = (u_0 \; e_0)^\top$,

$$A^t x_0 \approx (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \begin{pmatrix} u_0 \\ e_0 \end{pmatrix}$$

$$= (1+g)^t (u_0 + e_0) \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix} = n_t \bar{x},$$

where $n_t = (1+g)^t n_0$ and $n_0 = u_0 + e_0$

Regardless of x_0, state scales along $\bar{x}$ at rate of population growth

Rates

Unemployment rate $= u_t / n_t$

For large t, we have $u_t \approx n_t \bar{u}$

Hence unemployment rate $\approx (n_t \bar{u}) / n_t = \bar{u}$

Hence $\underline{\bar{u} \text{ is the long run rate of unemployment}}$

Similarly, $\bar{e}$ is the long run employment rate

$\implies$ dominant eigenvector gives unemployment rates

Extensions

Further analysis: how are α, λ, b and d determined?

For the hiring rate λ, we could use the job search model

In particular, with w^* as the reservation wage, we could set

$$\lambda = \mathbb{P}\{w_t \geqslant w^*\} = \sum_{w \in \mathsf{W}} \varphi(\mathbb{1}\{w \geqslant w^*\})$$

Doing so would allow us to study the crucial rate λ in terms of fundamental primitives, such as

- unemployment compensation
- impatience of individual agents, etc.

Markov Chains

Let

- $\mathsf{X} = \{x_1, \ldots, x_n\}$ = arbitrary finite set
- P be an $n \times n$ stochastic matrix

A Markov chain is generated by some stochastic matrix P

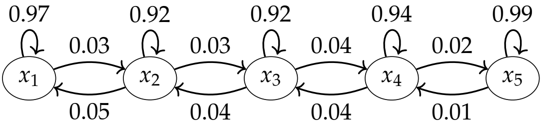
Interpretation:

$$P_{ij} = \text{ probability of moving from } x_i \text{ to } x_j \text{ in one step}$$

Example.

$$P = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix}$$

Transition probabilities:



Notation: We use the identification $P_{ij} :=: P(x_i, x_j)$

In this notation, P is a stochastic matrix iff

$$P \geqslant 0 \quad \text{and} \quad \sum_{x' \in \mathsf{X}} P(x, x') = 1 \text{ for all } x \in \mathsf{X}$$

Equivalent:

$$P \geqslant 0 \quad \text{and} \quad P\mathbb{1} = \mathbb{1}$$

Equivalent:

$$P(x, \cdot) \in \mathcal{D}(\mathsf{X}) \quad \text{for all } x \in \mathsf{X}$$

We call P "a stochastic matrix on X"

Let

- $(X_t)_{t \geqslant 0}$ be a sequence of X-valued random variables
- P be a stochastic matrix on X

Def. We call $(X_t)_{t \geqslant 0}$ P-**Markov** if

$$\mathbb{P}\{X_{t+1} = x' \mid X_0, X_1, \ldots, X_t\} = P(X_t, x') \quad \text{for all} \quad t \geqslant 0,\ x' \in \mathsf{X}$$

Standard terminology

- $(X_t)_{t \geqslant 0}$ is a **Markov chain**
- P is the **transition matrix** of $(X_t)_{t \geqslant 0}$
- We call either X_0 or its distribution ψ_0 the **initial condition**

Let

1. P be a stochastic matrix on X
2. ψ_0 be an element of $\mathcal{D}(\mathsf{X})$

This algorithm yields a P-Markov chain with initial condition ψ_0

$t \leftarrow 0$
$X_t \leftarrow$ a draw from ψ_0
while $t < \infty$ **do**
    $X_{t+1} \leftarrow$ a draw from the distribution $P(X_t, \cdot)$
    $t \leftarrow t + 1$
end

Application: Day laborer

A worker is either unemployed ($X_t = 1$) or employed ($X_t = 2$) each day

- In state 1 he is hired with probability $\alpha \in (0,1)$
- In state 2 he is fired with probability $\beta \in (0,1)$

The corresponding state space and transition matrix are

$$\mathsf{X} = \{1,2\} \quad \text{and} \quad P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

Ex. Show that

$$\psi^* := \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \end{pmatrix} = \text{ unique stationary distribution}$$

```julia
function create_laborer_model(; α=0.3, β=0.2)
    return (; α, β)
end

function laborer_update(x, model)  # update X from t to t+1
    (; α, β) = model
    if x == 1
        x′ = rand() < α ? 2 : 1
    else
        x′ = rand() < β ? 1 : 2
    end
    return x′
end
```

Ex. Write a function that simulates $(X_t)_{t \geqslant 0}$ given ψ_0, where

$$X_0 \sim \psi_0 = (p, 1 - p)$$

Using simulation, show that, for large k,

1. $\frac{1}{k} \sum_{t=1}^{k} \mathbb{1}\{X_t = 1\} \approx \psi^*(1)$
2. $\frac{1}{k} \sum_{t=1}^{k} \mathbb{1}\{X_t = 2\} \approx \psi^*(2)$

Check: this convergence does not depend on distribution of X_0

Below we explain why this is true

```
function sim_chain(k, p, model)
    X = Array{Int32}(undef, k)
    X[1] = rand() < p ? 1 : 2
    for t in 1:(k-1)
        X[t+1] = laborer_update(X[t], model)
    end
    return X
end

function test_convergence(; k=10_000_000, p=0.5)
    model = create_laborer_model()
    (; α, β) = model
    ψ_star = (1/(α + β)) * [β α]

    X = sim_chain(k, p, model)
    ψ_e = (1/k) * [sum(X .== 1)  sum(X .== 2)]
    error = maximum(abs.(ψ_star - ψ_e))
    approx_equal = isapprox(ψ_star, ψ_e, rtol=0.01)
    println("Sup norm deviation is $error")
    println("Approximate equality is $approx_equal")
 end
```

And now in Python

```python
import numpy as np
from collections import namedtuple
from numba import njit, int32

Model = namedtuple("Model", ("α", "β"), defaults=(0.3, 0.2))

@njit
def laborer_update(x, model):  # update X from t to t+1
    α, β = model
    if x == 1:
        y = 2 if np.random.rand() < α else 1
    else:
        y = 1 if np.random.rand() < β else 2
    return y
```

```python
@njit
def sim_chain(k, p, model):
    X = np.empty(k, dtype=int32)
    X[0] = 1 if np.random.rand() < p else 2
    for t in range(k-1):
        X[t+1] = laborer_update(X[t], model)
    return X

@njit
def test_convergence(model, k=10_000_000, p=0.5):
    α, β = model
    ψ_star = (1/(α + β)) * np.array((β,  α))

    X = sim_chain(k, p, model)
    ψ_e = (1/k) * np.array((sum(X == 1), sum(X == 2)))
    error = np.max(np.abs(ψ_star - ψ_e))
    return error

model = Model()
error = test_convergence(model)
print(f"Sup norm deviation is {error}")
```

Application: S-s Dynamics

Consider a firm whose inventory behavior follows S-s dynamics

Meaning:

- firm waits until its inventory falls below some level $s > 0$
- then replenishes by buying some fixed amount

Reasonable if ordering inventory involves a fixed cost

(We will see this behavior later in a DP problem with fixed costs)

Inventory $(X_t)_{t \geqslant 0}$ obeys

$$X_{t+1} = \max\{X_t - D_{t+1}, 0\} + S\mathbb{1}\{X_t \leqslant s\}, \tag{1}$$

where

- $(D_t)_{t \geqslant 1}$ is demand, IID with $D_t \stackrel{d}{=} \varphi \in \mathcal{D}(\mathbb{Z}_+)$ for all t and
- $S =$ amount of stock ordered when inventory $\leqslant s$

We assume φ obeys the geometric distribution:

$$\varphi(d) = \mathbb{P}\{D_t = d\} = p(1-p)^d \text{ for } d \in \mathbb{Z}_+$$

We take $\mathsf{X} := \{0, \dots, S + s\}$ to be the state space

Ex. Show that X satisfies

$$X_t \in \mathsf{X} \implies \mathbb{P}\{X_{t+1} \in \mathsf{X}\} = 1$$

Proof: Let $X_t = x \in S$, so that

$$X_{t+1} = \max\{x - D_{t+1}, 0\} + S\mathbb{1}\{x \leqslant s\}$$

Evidently $X_{t+1} \in \mathbb{Z}_+$. Also,

$$x \leqslant s \quad \implies \quad X_{t+1} = \max\{x - D_{t+1}, 0\} + S \leqslant s + S$$

and

$$s < x \leqslant S + s \quad \implies \quad X_{t+1} = \max\{x - D_{t+1}, 0\} \leqslant S + s$$

Ex. Show that X satisfies

$$X_t \in \mathsf{X} \implies \mathbb{P}\{X_{t+1} \in \mathsf{X}\} = 1$$

Proof: Let $X_t = x \in S$, so that

$$X_{t+1} = \max\{x - D_{t+1}, 0\} + S\mathbb{1}\{x \leqslant s\}$$

Evidently $X_{t+1} \in \mathbb{Z}_+$. Also,

$$x \leqslant s \quad \implies \quad X_{t+1} = \max\{x - D_{t+1}, 0\} + S \leqslant s + S$$

and

$$s < x \leqslant S + s \quad \implies \quad X_{t+1} = \max\{x - D_{t+1}, 0\} \leqslant S + s$$

If

$$h(x, d) = \max\{x - d, 0\} + S\mathbb{1}\{x \leqslant s\}$$

then

$$X_{t+1} = h(X_t, D_{t+1}) \quad \text{for all } t \geqslant 0$$

The transition matrix can be expressed as

$$P(x, x') = \mathbb{P}\{h(x, D_{t+1}) = x'\}$$

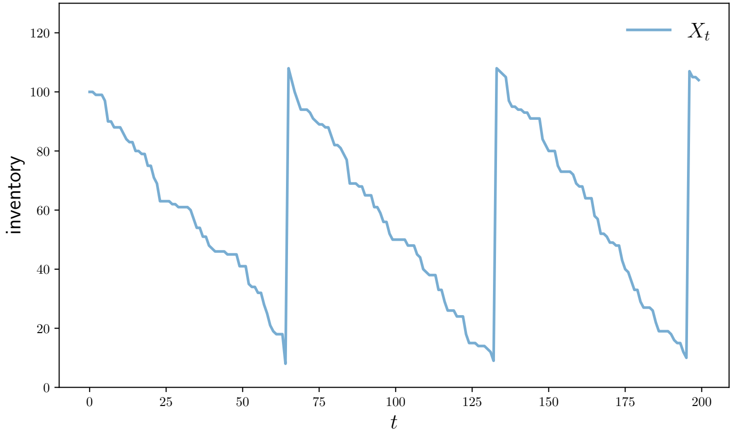
$$= \sum_{d \geqslant 0} \mathbb{1}\{h(x, d) = x'\}\varphi(d)$$

(In calculations we truncate the sum)

```julia
using Distributions, IterTools, QuantEcon

function create_inventory_model(; S=100,  # Order size
                                  s=10,   # Order threshold
                                  p=0.4)  # Demand parameter
    ϕ = Geometric(p)
    h(x, d) = max(x - d, 0) + S*(x <= s)
    return (; S, s, p, ϕ, h)
end

"Simulate the inventory process."
function sim_inventories(model; ts_length=200)
    (; S, s, p, ϕ, h) = model
    X = Vector{Int32}(undef, ts_length)
    X[1] = S  # Initial condition
    for t in 1:(ts_length-1)
        X[t+1] = h(X[t], rand(ϕ))
    end
    return X
end
```



X_t
inventory
t
0
20
40
60
80
100
120
0
25
50
75
100
125
150
175
200

Multistep transitions

Fix a state space X and transition matrix P on X

The k-th power P^k is called the **k-step transition matrix**

- $P^k(x, x')$ denotes the (x, x')-th element of P^k

<u>Claim</u>:

$$P^k(x, x') = \mathbb{P}\{X_{t+k} = x' \mid X_t = x\} \quad \text{for any } P\text{-chain } (X_t)_{t \geqslant 0}$$

This claim can be verified by induction

Fix $t \in \mathbb{N}$ and $x, x' \in$ X

True by definition when $k = 1$

Multistep transitions

Fix a state space X and transition matrix P on X

The k-th power P^k is called the **k-step transition matrix**

- $P^k(x, x')$ denotes the (x, x')-th element of P^k

Claim:

$$P^k(x, x') = \mathbb{P}\{X_{t+k} = x' \mid X_t = x\} \quad \text{for any } P\text{-chain } (X_t)_{t \geqslant 0}$$

This claim can be verified by induction

Fix $t \in \mathbb{N}$ and $x, x' \in$ X

True by definition when $k = 1$

Now suppose true at k, so that

$$P^k(x, x') = \mathbb{P}\{X_{t+k} = x' \mid X_t = x\}$$

By the law of total probability, we have

$$\mathbb{P}\{X_{t+k+1} = x' \mid X_t = x\}$$

$$= \sum_z \mathbb{P}\{X_{t+k+1} = x' \mid X_{t+k} = z\} \mathbb{P}\{X_{t+k} = z \mid X_t = x\}$$

Applying the induction hypothesis, the last equation becomes

$$\mathbb{P}\{X_{t+k+1} = x' \mid X_t = x\} = \sum_z P^k(x, z) P(z, x') = P^{k+1}(x, x')$$

This completes our proof by induction

Lemma. The following statements are equivalent:

1. P is irreducible
2. For any P-chain (X_t) and any $x, x' \in \mathsf{X}$, there exists a $k \geqslant 0$ such that
$$\mathbb{P}\{X_k = x' \mid X_0 = x\} > 0$$

Proof:

$$\text{statement 1} \iff \sum_{k \geqslant 0} P^k \gg 0$$

$$\iff \forall\, x, x' \in \mathsf{X},\ \exists\, k \geqslant 0 \text{ s.t. } P^k(x, x') > 0$$

$$\iff \text{statement 2}$$

Dynamics of Marginals

Fix a stochastic matrix P on X and let (X_t) be a P-chain

By the law of total probability, for all $x, x' \in \mathsf{X}$,

$$\mathbb{P}\{X_{t+1} = x'\} = \sum_{x \in \mathsf{X}} \mathbb{P}\{X_{t+1} = x' \mid X_t = x\} \mathbb{P}\{X_t = x\}$$

Let $\psi_t :\stackrel{d}{=} X_t$ for all t

Using this notation, we can rewrite the last display as

$$\psi_{t+1}(x') = \sum_{x \in \mathsf{X}} P(x, x') \psi_t(x) \quad \text{for all } x \in \mathsf{X}$$

Treating each ψ_t as a <u>row</u> vector, we get $\psi_{t+1} = \psi_t P$

Stationarity

Distributions update via $\psi_{t+1} = \psi_t P$

Recall also that ψ^* is called **stationary** for P if $\psi^* = \psi^* P$

Now we can interpret this expression

If ψ^* is stationary for P, then

$$X_t \stackrel{d}{=} \psi^* \implies X_{t+1} \stackrel{d}{=} \psi^*$$

We recall that

- every stochastic matrix on X has at least one stationary distribution, and
- uniqueness in $\mathcal{D}(\mathsf{X})$ holds whenever P is irreducible

Ergodicity

Theorem. Let P be irreducible with stationary distribution ψ^*

For any P-Markov chain (X_t) and any $x \in \mathsf{X}$, we have

$$\mathbb{P}\left\{\lim_{k\to\infty}\frac{1}{k}\sum_{t=0}^{k-1}\mathbb{1}\{X_t = x\} = \psi^*(x)\right\} = 1$$

Meaning: For (almost) every P-Markov chain that we generate,

fraction of time chain spends in state $x \approx \psi^*(x)$

Markov chains with this property are sometimes said to be **ergodic**

Example. Recall the model

$$\mathsf{X} = \{1, 2\}, \quad P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \quad \text{and} \quad \psi^* := \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \end{pmatrix}$$

Since P is irreducible, ergodicity holds:

$$\mathbb{P}\left\{ \lim_{k \to \infty} \frac{1}{k} \sum_{t=0}^{k-1} \mathbb{1}\{X_t = x\} = \psi^*(x) \right\} = 1$$

This is what we saw in the simulations above

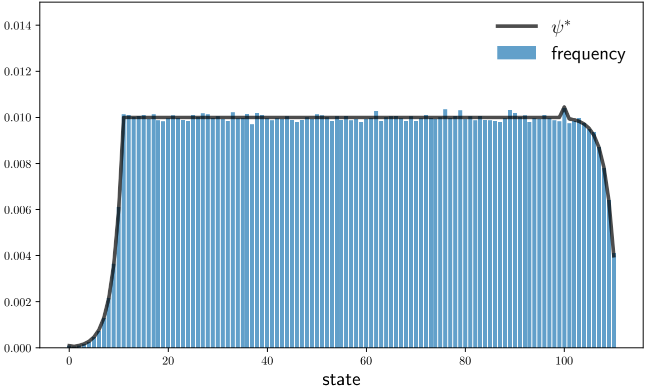


Figure: Ergodicity in the inventory model

Ex. Prove $P \gg 0$ implies $\psi P^t \to \psi^*$ as $t \to \infty$ for any $\psi \in \mathcal{D}(\mathsf{X})$

Proof: P is positive and $r(P) = 1$

PF theorem implies $P^t \to e\,\varepsilon$ as $t \to \infty$, where $\langle e, \varepsilon \rangle = 1$

In this case we know

- $\mathbb{1}$ is the dominant right eigenvector,
- ψ^* is the dominant left eigenvector and $\langle \psi^*, \mathbb{1} \rangle = \psi^* \mathbb{1} = 1$

Hence, for any $\psi \in \mathcal{D}(\mathsf{X})$, we have

$$\psi P^t \to \psi \mathbb{1}\, \psi^* = \psi^* \quad \text{as} \quad t \to \infty$$

Ex. Prove $P \gg 0$ implies $\psi P^t \to \psi^*$ as $t \to \infty$ for any $\psi \in \mathcal{D}(\mathsf{X})$

Proof: P is positive and $r(P) = 1$

PF theorem implies $P^t \to e\,\varepsilon$ as $t \to \infty$, where $\langle e, \varepsilon \rangle = 1$

In this case we know

- $\mathbb{1}$ is the dominant right eigenvector,
- ψ^* is the dominant left eigenvector and $\langle \psi^*, \mathbb{1} \rangle = \psi^* \mathbb{1} = 1$

Hence, for any $\psi \in \mathcal{D}(\mathsf{X})$, we have

$$\psi P^t \to \psi \mathbb{1}\, \psi^* = \psi^* \quad \text{as} \quad t \to \infty$$

Recall the model

$$\mathsf{X} = \{1,2\}, \quad P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \quad \text{and} \quad \psi^* := \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \end{pmatrix}$$

Ex. Fix $\alpha = 0.3$ and $\beta = 0.2$

Compute the sequence (ψP^t) for different choices of ψ

Confirm that your results are consistent with the claim that

$$\psi P^t \to \psi^* \text{ as } t \to \infty \text{ for any } \psi \in \mathcal{D}(\mathsf{X})$$

Approximation

It can be helpful to reduce continuous state Markov models to finite state models

For example, suppose that $(X_t)_{t \geqslant 0}$ evolves in $\mathbb{R}$ according to

$$X_{t+1} = \rho X_t + b + \nu \varepsilon_{t+1}, \quad (\varepsilon_t) \overset{\text{IID}}{\sim} N(0,1). \tag{2}$$

This is a **linear Gaussian AR(1)** model

To approximate it we use **Tauchen's method**

We assume throughout that $|\rho| < 1$

Under this assumption, (2) has a unique **stationary distribution** ψ^* given by

$$\psi^* = N(\mu_x, \sigma_x^2) \quad \text{with} \quad \mu_x := \frac{b}{1-\rho} \quad \text{and} \quad \sigma_x^2 := \frac{\nu^2}{1-\rho^2}$$

This means that ψ^* has the following property:

$$X_t \stackrel{d}{=} \psi^* \text{ and } X_{t+1} = \rho X_t + b + \nu \varepsilon_{t+1} \text{ implies } X_{t+1} \stackrel{d}{=} \psi^*$$

Ex. Prove this. Hints: When $X_t \stackrel{d}{=} \psi^*$,

- is X_{t+1} normally distributed?
- what is its mean and variance?

Tauchen's discretization method

We start with the case $b = 0$

Fix $m, n \in \mathbb{N}$

Create state space $\mathsf{X} := \{x_1, \ldots, x_n\}$ via

- set $x_1 = -m\,\sigma_x$,
- set $x_n = m\,\sigma_x$ and
- set $x_{i+1} = x_i + s$ for $i \in \{1, \ldots, n-1\}$ where

$$s = \frac{x_n - x_1}{n - 1}$$

A grid that brackets the stationary mean on both sides by m standard deviations:

Create an $n \times n$ matrix P such that, For $i, j \in [n]$,

1. if $j = 1$, then set $P(x_i, x_j) = F(x_1 - \rho x_i + s/2)$
2. If $j = n$, then set $P(x_i, x_j) = 1 - F(x_n - \rho x_i - s/2)$
3. Otherwise, set

$$P(x_i, x_j) = F(x_j - \rho x_i + s/2) - F(x_j - \rho x_i - s/2)$$

Finally, if $b \neq 0$, then replace x_i with $x_i + \mu_x$ for each i

- shift the grid X to be centered on the stationary mean

```julia
using QuantEcon

ρ, b, ν = 0.9, 0.0, 1.0
μ_x = b/(1-ρ)
σ_x = sqrt(ν^2 / (1-ρ^2))

n = 15
mc = tauchen(n, ρ, ν)
approx_sd = stationary_distributions(mc)[1]

function psi_star(y)
    c = 1/(sqrt(2 * pi) * σ_x)
    return c * exp(-(y - μ_x)^2 / (2 * σ_x^2))
end
```

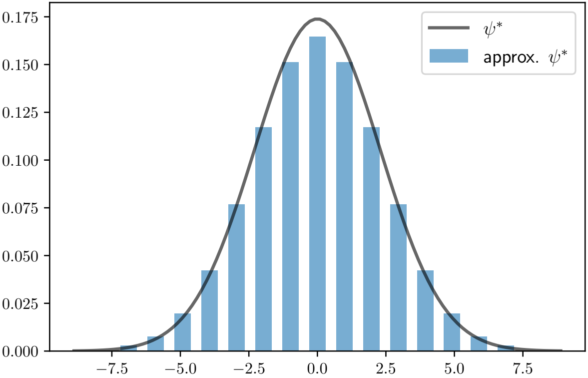


Figure: Comparison of $\psi^* = N(\mu_x, \sigma_x^2)$ and its discrete approximant

Conditional Expectations

Let P be any stochastic matrix on X

For each $h \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$, we define

$$(Ph)(x) := \sum_{x' \in \mathsf{X}} h(x') P(x, x')$$

Equivalently

$$(Ph)(x) = \mathbb{E}[h(X_{t+1}) \mid X_t = x] \quad \text{when } (X_t) \text{ is } P\text{-Markov}$$

This interpretation extends to powers:

$$(P^k h)(x) = \sum_{x' \in \mathsf{X}} h(x') P^k(x, x') = \mathbb{E}[h(X_{t+k}) \mid X_t = x]$$

Quick note on **conventions**

When updating distributions we use row vectors:

$$(\psi_t P)(x') = \sum_{x \in X} P(x, x') \psi_t(x)$$

When taking conditional expectations we use column vectors:

$$(Ph)(x) := \sum_{x' \in X} h(x') P(x, x')$$

The Law of Iterated Expectations

We now prove a version of the **law of iterated expectations**

Let (X_t) be P-Markov with $X_0 \stackrel{d}{=} \psi_0$

Fix $t, k \in \mathbb{N}$ and set $\mathbb{E}_t := \mathbb{E}[\cdot \mid X_t]$

We claim that

$$\mathbb{E}[\mathbb{E}_t[h(X_{t+k})]] = \mathbb{E}[h(X_{t+k})] \quad \text{for any } h \in \mathbb{R}^{\mathsf{X}}$$

(A special case of the general rule $\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$)

To see this, recall that $\mathbb{E}[h(X_{t+k}) \mid X_t = x] = (P^k h)(x)$

Hence $\mathbb{E}[h(X_{t+k}) \mid X_t] = (P^k h)(X_t)$

Therefore,

$$\mathbb{E}[\mathbb{E}_t[h(X_{t+k})]] = \mathbb{E}[(P^k h)(X_t)]$$

$$= \sum_{x'} (P^k h)(x') \psi_t(x') = \sum_{x'} (P^k h)(x') (\psi_0 P^t)(x')$$

Since $\psi_0 P^t$ is a row vector, we can write the last expression as

$$\psi_0 P^t P^k h = \psi_0 P^{t+k} h = \psi_{t+k} h = \mathbb{E} h(X_{t+k})$$