Modeling and Analysis of Markov Processes with Non-stationary Transition Probabilities

Peter W. Glynn

Stanford University
Institute for Advanced Study, City University of Hong Kong

IAS Distinguished Lecture Series Frontiers in Operations Research/Operations Management
October 29th, 2018

Outline of Talk

1. What does non-stationary modeling mean?
2. Why is non-stationary modeling important?
3. Comparison with stationary modeling
4. Approximations for non-stationary systems
5. Fast numerical solvers for non-stationary systems
6. Simulation algorithms for non-stationary systems

1. What does non-stationary modeling mean?

All real-world systems arising in service operations, logistics, and production settings exhibit:

- time-of-day effects
- day-of-week effects
- seasonality effects
- secular trends

Modeling with stationary transition probabilities assumes that the dynamics are independent of time:

- constant arrival rate
- constant demand distribution
- constant abandonment rate

Models with stationary transition dynamics

- $X_n = f(X_{n-1}, Z_n)$

  ↖ iid "noise"

- Markov chain with stationary transition probabilities

$$P(X_n = y | X_{n-1} = x) \triangleq P(x, y)$$

- Dynamics characterized by one-step transition matrix

$$P = (P(x, y) : x, y \in S)$$

The vast majority of stochastic modeling papers focus on such "stationary" models

$$\downarrow$$

We will discuss why this is the case later

In this lecture, we will discuss general tools for dealing with stochastic systems with *non-stationary* dynamics:

- $X_n = f_n(X_{n-1}, Z_n)$

  $\nwarrow$ independent "noise"

- Markov chain with non-stationary transition probabilities

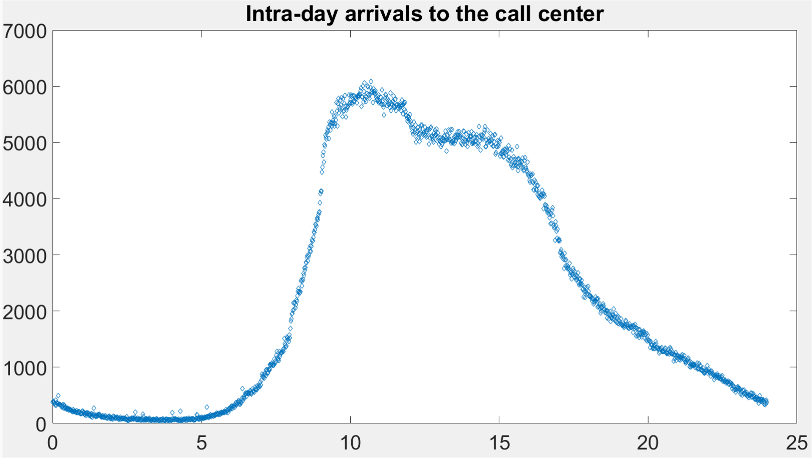
$$P(X_n = y | X_{n-1} = x) = P_n(x, y)$$

- Dynamics characterized by a sequence of transition matrices:

$$P_1, P_2, P_3, \ldots$$

  where $P_i = (P_i(x, y) : x, y \in S)$

2. Why is non-stationary modeling important?



These non-stationarities have always been present in the real-world...

Why so little research to date on this class of models?

Stationary models are much easier to analyze mathematically (e.g. steady-state behavior)

and

much easier to interpret (e.g. optimal policies are much simpler to describe)

Trade-off tractability against model fidelity

Appropriate when model is being used “descriptively”

Descriptive modeling has perhaps been the principal driver for the use of OR-based stochastics historically

- Used to generate insight into complex engineering or managerial problems
    - Design of communications networks, scheduling rules, etc
- Many important structural insights have been obtained
- Can help identify the right class of policies to consider

But stochastic modeling has a major role to play in a world in which

predictive analysis

and

prescriptive analysis

will become increasingly important

In such a world, the need for high-fidelity models increases

$\downarrow$

Incorporating non-stationarities becomes more important

Comparison with Stationary Modeling

For Markov chains with stationary transition probabilities, the *n*-step transition probabilities

$$P_n(x, y) = P(X_n = y | X_0 = x)$$

satisfy:

$$P_n = P^n$$

where $P_n = (P_n(x, y) : x, y \in S)$

Result: If $X = (X_n : n \geq 0)$ is an irreducible aperiodic positive recurrent Markov chain, then

$$P_n(x, y) \to \pi(y)$$

as $n \to \infty$, where $\pi = (\pi(y) : y \in S)$ is the unique probability mass function satisfying

$$\pi = \pi P.$$

π is called the

- steady-state distribution
- stationary distribution
- invariant distribution
- equilibrium distribution

Computing π in closed form is possible for many Markov chain models:

- birth-death chains
- Jackson/Kelly networks
- product-form networks
- etc

Computing π is the central focus of stationary Markov chain modeling

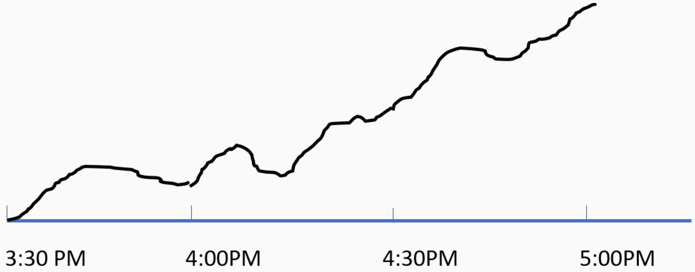
Note that

$$P(X_n = y | X_0 = x) \approx \pi(y)$$

is approximately independent of x

"loss of memory"

For Markov chains with non-stationary transition probabilities, no steady-state typically exists



Queue with slowly increasing arrival rate

An important exception:

- Non-stationary periodic Markov chains

$$P_{i+p} = P_i, \quad i \geq 0$$

-

$$\begin{aligned} p = & 24, \quad 1 \text{ day} \\ & 168, \quad 1 \text{ week} \\ & \text{etc} \end{aligned}$$

For such chains,

$$P_{np+i}(x, y) \to \pi_i(y)$$

as $n \to \infty$, where

$$\pi_i = \pi_0 P_1 \cdots P_{i-1}, \quad 0 \leq i \leq p$$

and

$$\pi_0 = \pi_0 R$$

with

$$R = P_1 P_2 \cdots P_p$$

"periodic equilibrium"

- Typical of

  many service operations settings

  traffic congestion in urban areas

  etc

- Note that $\pi_0 = \pi_0 R$ involves

$$R = P_1 P_2 \cdots P_p$$

- R will almost never be available in closed form

Almost nothing in the non-stationary setting is computable in closed-form

$$\downarrow$$

Focus on:

- fast computational algorithms
- approximations

Notable exception:

Non-stationary infinite-server queues can be easily analyzed

4. Approximations for non-stationary Markov chains

Our philosophy:

- approximations that do not rely on model structure

  very general

- require numerical computation comparable to stationary case
- are asymptotically valid in regime that arises naturally in practice
- can be easily implemented

Existing approaches:

- very model-specific (e.g., $M/M/1$; $M/M/s$)
- can be difficult to implement in practice (e.g., no guidelines on how to set parameters)

The Pointwise Stationary Approximation (PSA)

To approximate the behavior of a non-stationary Markov chain $X = (X_k : k \geq 0)$ t time n, use

$$\mathbb{E}r(X_n) \approx \sum_x \pi_n(x) r(x) = \pi_n r$$

where π_n is the stationary distribution of P_n:

$$\pi_n = \pi_n P_n$$

Remarks:

- Most widely used general purpose approximation for $\mathbb{E}r(X_n)$
- Valid when the P_j's are essentially constant for j "close to" n

Why it works:

- Let $\mu = (\mu(x) : x \in S)$ be the "initial distribution" for X_0
- Then,

$$\mathbb{E}r(X_n) = \mu P_1 P_2 \cdots P_n r$$

- If $P_{n-k} \approx P_n$ for $0 \leq k \leq m$, then

$$\mu P_1 \cdots P_n r \approx \mu P_1 \cdots P_{n-m-1} P_n^m r \approx \mu P_1 \cdots P_{n-m-1} \Pi_n r = \pi_n r$$

We want to improve upon PSA

Assume that P_j's change slowly

e.g. $P_k = P(kh)$ where $P(\cdot)$ is "smooth"

i.e. $P_{n-k} = P((n-k)h) \approx P(nh) - khP'(nh) + O(h^2)$

Step 1:

$$\mu P_1 \cdots P_n r = \mu(P_1 \cdots P_{n-m-1})(P_{n-m} \cdots P_n) r$$

and

$$\begin{aligned}
P_{n-m} \cdots P_n &\approx (P(nh) - mhP'(nh)) \cdots (P(nh) - hP'(nh))P(nh) \\
&\approx P(nh)^{m+1} - h \sum_{j=1}^{m} jP(nh)^{m-j} P'(nh) P(nh)^j \\
&\approx \Pi(nh) - h \sum_{j=1}^{\infty} j\Pi(nh) P'(nh) P(nh)^j \\
&= \Pi(nh) - h\Pi(nh) \sum_{j=1}^{\infty} jP'(nh) P(nh)^j
\end{aligned}$$

Step 2:

$$\frac{d}{dt}\sum_y P(nh+t,x,y)=0$$

so

$$P'(nh)P(nh)^j = P'(nh)(P(nh)^j - \Pi(nh))$$

Also,

$$\begin{aligned}(P(nh)-\Pi(nh))^2 &= P^2(nh) - P(nh)\Pi(nh) - \Pi(nh)P(nh) + \Pi(nh)^2\\ &= P^2(nh) - \Pi(nh)\end{aligned}$$

so

$$\sum_{j=1}^{\infty} jP'(nh)P(nh)^j = P'(nh)\sum_{j=1}^{\infty} j(P(nh)-\Pi(nh))^j$$

Step 3:

$$\Big(\sum_{j=0}^{\infty} A^j\Big)\Big(\sum_{j=0}^{\infty} A^j\Big) = \sum_{j=0}^{\infty}(j+1)A^j$$

so

$$\sum_{j=0}^{\infty} jA^j = (I-A)^{-2} - (I-A)^{-1}$$

Put it all together...

$$\mu P_1 P_2 \cdots P_n r = \pi_n r - h\pi_n P_n'((I - P_n + \Pi_n)^{-2} - (I - P_n + \Pi_n)^{-1})r$$

Note that

$$hP_n' \approx P_n - P_{n-1}$$

$$\boxed{\mathbb{E}r(X_n) \approx \pi_n r - \pi_n(P_n - P_{n-1})((I - P_n + \Pi_n)^{-2} - (I - P_n + \Pi_n)^{-1})r}$$

An improved approximation to PSA Zheng, Honnappa, G (2018)

Approximation involves:

- Computing solution to $\pi_n = \pi_n P_n$
- Solving Poisson's equation twice:

$$(I - P_n)g_n = r - \Pi_n r$$
$$(I - P_n)h_n = g_n - \Pi_n g_n$$

- We can similarly develop higher-order approximations
- related to "uniform acceleration expansion" Massey and Whitt (1998)

Can also be carried out for

- discounted reward

$$\sum_{j=0}^{\infty} e^{-\alpha j} \mathbb{E} r(X_j)$$

- expected hitting time

$$\mathbb{E} T$$

- etc

whenever the Markov chain has slowly changing transition probabilities

Linear system for time-varying “correction” is always of the same form as linear system in the stationary case

e.g. discounted reward

$$(I - e^{-\alpha} P_n) g = f$$

Extends to:

- Markov jump processes
- Diffusions
- Reflected Brownian motion

etc

Fast Numerical Solvers for Non-stationary Systems (with A. Infanger)

- In many settings, approximations will not be good enough
- Need fast numerical algorithms for computations

$$\mathbb{E}r(X_n) = \mu P_1 P_1 \cdots P_n r$$

Periodic systems:

Algorithm 1.

- Compute

$$\mu P_1 P_2 \cdots P_j$$

  recursively in j until $j = n$

- Complexity: $O(n|S|^2)$

Algorithm 2.

- First compute

$$R = P_1 P_2 \cdots P_p$$

Complexity: $O(p|S|^3)$

- If $n = mp + k$, then compute

$$\mu R^m P_1 P_2 \cdots P_{k-1}$$

Complexity: $O((m+k)|S|^2)$

Algorithm 1 is better than Algorithm 2 if $|S| >> m$.

Algorithm 3.

- Compute backwards recursively in k

$$u_k \triangleq P_{n-k} P_{n-k+1} \cdots P_n r$$

  until

$$\sup_{x,y} |u_k(x) - u_k(y)| < \epsilon$$

- Then,

$$\|u_k - u_n\| < \epsilon$$

Why this works:

- Non-stationary chains also exhibit “loss of memory”

$$(P_{n-k} \cdots P_n)(x, y)$$

  is independent of x for k large enough

- So, what happens to the chain over $[0, n-k]$ is irrelevant

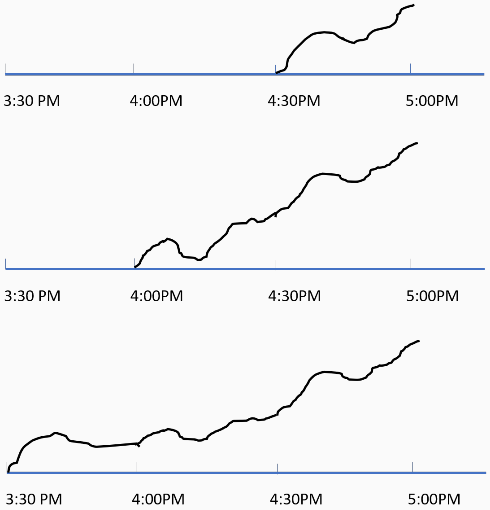
Goal: Analyze freeway congestion at 5 PM

- Simulate vehicle traffic

- Start with freeway empty at 4:30 PM?
  4 PM?
  3:30 PM?

- How far back?

Backward Coupling



- $X_{n,k} = \varphi_k(Z_{n-k-1}, \ldots, Z_n)$
- when

$$X_{n,k} \approx X_{n,k+1} \ ,$$

  then k is large enough

Zheng and G (2019)

Concluding Remarks

- Non-stationary models will become more important as OR/MS moves toward more predictive/prescriptive stochastic modeling

- Refined approximations for slowly changing Markov chains are broadly applicable

- Fast algorithms for non-stationary models are becoming available based on

  numerical linear algebra
  simulation/Monte Carlo