دانشکده مهندسی کامپیوتر

عنوان درس:

ارزیابی کارایی سیستم‌های کامپیوتری

Performance Evaluation of Computer Systems (PECS)

جلسه ۹: زنجیره‌های مارکوف

مدرس:

محمد عبداللهی ازگمی

(Mohammad Abdollahi Azgomi)

azgomi@iust.ac.ir

زنجیره‌های مارکوف

- مقدمه
- فرآیند مارکوف
- زنجیره مارکوف
- زنجیره‌های مارکوف گسسته-زمان (DTMCs)
  - حل حالت پایدار DTMCs
  - حل حالت گذرای DTMCs
  - مثالهایی از DTMCs
- زنجیره‌های مارکوف پیوسته-زمان (CTMCs)
  - حل حالت پایدار CTMCs
  - حل حالت گذرای CTMCs
  - مثالهایی از CTMCs
- حل عددی زنجیره‌های مارکوف پیوسته-زمان
- جمع‌بندی

مقدمه

- نظریه فرآیندهای مارکوف منتسب به ریاضیدان روسی به نام **آندری آندریویچ مارکوف** (Andrey Andreyevich Markov) (۱۸۵۶-۱۹۲۲) است که از افراد پیشتاز در تحقیقات سیستماتیک توصیف ریاضی فرآیندهای تصادفی است.
- فرآیند مارکوف (Markov process) ابزاری بسیار منعطف، قوی و کاراً برای توصیف و تحلیل خصوصیت‌های سیستمهای پویای کامپیوتری و غیرکامپیوتری هستند.
- از مزایای این مدلها در مورد سیستمهای کامپیوتری، محاسبه و استخراج آسان معیارهای کارآیی و اتکاءپذیری از آنها است.

مقدمه

- فرآیند مارکوف نظریه مبنایی حل سیستم‌های صف است.
- در حقیقت، نماد سیستمهای صف خیلی از اوقات به‌عنوان یک فن توصیف سطح بالا برای برخی از انواع فرآیندهای مارکوف در نظر گرفته می‌شود.
- هر سیستم صف در اصل می‌تواند به یک نمونه از فرآیند مارکوف نگاشت شده و آنگاه با روشهای ریاضی بر اساس آن فرآیند ارزیابی شود.
- اغلب خصوصیت‌های سیستمهای صف بر اساس فرآیندهای مارکوف اثبات می‌شود.
- متدولوژی فوق حتی برای آن دسته از سیستمهای صف که مارکوفی نیستند نیز قابل بهره‌برداری است. برای این منظور از روشهای مارکوف‌سازی (Markovizing) یا متغیرهای متمم (suplementary variables) برای این منظور قابل استفاده هستند.

فرآیند مارکوف

- یک نوع خاص از فرآیندهای تصادفی که مورد بررسی دقیق‌تر قرار می‌دهیم، **فرآیند مارکوف** (Markov process) نامیده می‌شود.
- یک فرآیند مارکوف به‌طور غیرصوری به‌شکل زیر تعریف می‌شود:
  - با داشتن حالت (state) یا مقدار (value) یک فرآیند مارکوف X در زمان t (یعنی X(t))، رفتار آینده X به‌طور کامل بر حسب X(t) قابل توصیف خواهد بود.
- **فرآیند مارکوف** دارای این خاصیت سودمند است که رفتار آینده آن مستقل از مقادیر گذشته آن است.
  - یعنی بی‌حافظه است.

فرآیند مارکوف (ادامه)

- به بیان دقیق و صوری، فرآیند مارکوف به‌صورت زیر تعریف می‌شود:
  - فرآیند {X(t)} دارای خاصیت مارکوفی (Markovian property) یا بی‌حافظگی است اگر با داشتن مقدار X(t) در زمان $t \in T$، مسیر آینده X(s) برای $s > t$ به دانستن اینکه سابقه گذشته X(u) برای $u < t$ چه بوده است بستگی نداشته باشد. یعنی برای $t_1 < \ldots < t_n < t_{n+1}$ داشته باشیم:

$$P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n, \ldots, X(t_1) = x_1] = P[X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n]$$

زنجیره مارکوف

- یک **زنجیره مارکوف** (Markov chain)، یک فرآیند مارکوف دارای فضای حالت گسسته است.
- ما همیشه فرض می‌کنیم که یک زنجیره مارکوف دارای فضای حالتی در مجموعه {0, 1, 2, ...} است و نیز **زمان – همگن** (time-homogeneous) است.
- همگن‌بودن زمان زنجیره مارکوف به‌صورت زیر تعریف می‌شود:
  - رفتار سیستم به اینکه چه موقع سیستم مشاهده شده است وابسته نباشد.
  - به عبارت دقیق تر، نرخهای گذر (transition rates) بین حالتها مستقل از زمانی که گذر در آن اتفاق می افتد باشد. بنا بر این برای همه tها و sها داشته باشیم:

$$P[X(t+k) = x_k \mid X(t) = x_j] = P[X(s+k) = x_k \mid X(s) = x_j]$$

زنجیره‌های مارکوف گسسته-زمان

- یک **زنجیره مارکوف گسسته-زمان** (DTMC: discrete-time M.C.) است اگر:

$$P\left[X(t+k)=j \middle| X(t)=i, X(t-1)=n_{t-1}, X(t-2)=n_{t-2}, \ldots, X(O)=n_O\right]$$
$$= P\left[X(t+k)=j \middle| X(t)=i\right] \quad (1)$$
$$= P_{ij}^{(k)} \quad (2)$$

  - توجه کنید که به ازای i و j داده شده، $P_{ij}^{(k)}$ یک عدد است.
  - $P_{ij}^{(k)}$ می‌تواند به‌عنوان احتمال اینکه اگر X دارای مقدار i است بعد از k *گام زمانی* (time-step) دارای مقدار j بشود، توصیف شود.
  - اغلب از P_{ij} به‌جای $P_{ij}^{(1)}$ استفاده می‌شود: *احتمال یک گامی* (one-step probability).

ماتریس احتمال گذر

■ در حقیقت اگر ما همه P_{ij}ها را در نظر بگیریم، یک ماتریس به شکل $P = \{p_{ij}\}$ خواهیم داشته که مشخصه هر زنجیره مارکوف بوده و به آن **ماتریس احتمال گذر** (TPM: transition probability matrix) می‌گوییم، که یک ماتریس nxn است:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

بردار احتمال اشغال حالت

■ حال **بردار احتمال اشغال حالت** (state occupancy probability vector) را تعریف می‌کنیم:

- فرض کنید که π یک بردار ردیفی (row vector) باشد. عنصر i–ام این بردار را با π_i نشان می‌دهیم:

$$\pi = [\pi_1, \pi_2, \pi_3, \ldots \pi_{|S|}]$$

- اگر π یک بردار اشغال حالت باشد، آنگاه $\pi_i(k)$ احتمال آن خواهد بود که DTMC بعد از k گام زمانی دارای مقدار i بشود (یا در حالت i باشد).
- فرض کنید که یک DTMC که با X نشان می‌دهیم دارای فضای حالت به اندازه n است (یعنی $S = \{1, 2, \ldots, n\}$). به بیان صوری می‌گوییم که:

$$\sum_{i=1}^{n} \pi_i(k) = 1$$ (k) = i]

- توجه کنید که برای همه زمانهای k داریم:

محاسبه بردار احتمال اشغال حالت

- محاسبه بردار احتمال اشغال حالت یک گام مجزا به جلو در زمان (a single step forward in time) در زمان است.
- اگر $\pi(0)$ (بردار احتمال اولیه) را داشته باشیم و P_{ij} برای i, j = 1, 2, …, n، چگونه می‌توانیم $\pi(1)$ را محاسبه کنیم؟
  - ابتدا تعریف P_{ij} را به یاد بیاورید:

$$\begin{aligned} P_{ij} &= P[X(k+1) = j \mid X(k) = i] \\ &= P[X(1) = j \mid X(0) = i] \end{aligned}$$

  - چون داریم: $\sum_{i=1}^{n} \pi_i(0) = 1$
  - آنگاه با توجه به $\pi_i(k) = P[X(k) = i]$ و تعریف DTMC خواهیم داشت:

$$\begin{aligned} \pi_j(1) &= P[X(1) = j] \\ &= P[X(1) = j|X(0) = 1]P[X(0) = 1] + \ldots + P[X(1) = j|X(0) = n]P[X(0) = n] \\ &= \sum_{i=1}^{n} P[X(1) = j|X(0) = i]P[X(0) = i] \\ &= \sum_{i=1}^{n} P_{ij}\pi_i(0) \\ &= \sum_{i=1}^{n} \pi_i(0)P_{ij} \end{aligned}$$

The law of total probability:

$$\boxed{P(A) = \sum_{i=1}^{N} P(A_i)\, P(A \mid A_i),}$$

محاسبه بردار احتمال اشغال حالت

- بنابر این با توجه به آنچه که دیدیم داریم:

$$\pi_j(1) = \sum_{i=1}^{n} \pi_i(0)P_{ij}$$

که برای همه مقادیر j صادق است و محاسبه فوق مشابه ضرب بردار در ماتریس است و خواهیم داشت:

$$\pi(1) = \pi(0)\, P$$

  - که در آن $\pi(0)$ و $\pi(1)$ بردارهای ردیفی هستند و $\pi(0)\, P$ یک ضرب بردار در ماتریس است.
- نتیجه مهمی که حاصل می‌شود این است که ما می‌توانیم به آسانی یک DTMC را بر حسب یک بردار احتمال اشغال حالت (π) و یک ماتریس احتمال گذر (P) مشخص کنیم.

رفتار حالت گذاری زنجیره‌های مارکوف گسسته-زمان

- با داشتن $\pi(0)$ و P، چطور می‌توان $\pi(k)$ را محاسبه نمود؟
- حالت عمومی رابطه قبلی به‌صورت زیر است:

$$\pi(k) = \pi(k-1)P$$

- همچنین، چون $\pi(k-1) = \pi(k-2)P$ می‌توانیم بنویسیم:

$$\pi(k) = (\pi(k-2)P)P = \pi(k-2)P^2$$

- به‌طور مشابه، چون $\pi(k-2) = \pi(k-3)P$ خواهیم داشت:

$$\pi(k) = [(\pi(k-3)P]P^2 = \pi(k-3)P^3$$

- با تکرار عمل فوق به‌آسانی می‌توان دید که داریم:

$$\pi(k) = \pi(0)P^k$$

یک مثال از زنجیره‌های مارکوف گسسته-زمان

- فرض کنید که هوای یک شهر به‌صورت زیر مدل شده است:
  - اگر امروز هوا آفتابی است، با احتمال ۰.۶ فردا نیز آفتابی خواهد بود، با احتمال ۰.۳ ابری و با احتمال ۰.۱ بارانی خواهد بود.
  - اگر امروز هوا ابری است، با احتمال ۰.۴ فردا آفتابی، با احتمال ۰.۴۵ فردا ابری و با احتمال ۰.۱۵ فردا بارانی خواهد بود.
  - اگر امروز بارانی است، با احتمال ۰.۱۵ فردا آفتابی، با احتمال ۰.۶ فردا ابری و با احتمال ۰.۲۵ فردا بارانی خواهد بود.
- حال اگر جمعه بارانی باشد، پیش‌بینی هوا برای دوشنبه چه خواهد بود؟
- مدل وضع هوا چه نوع زنجیره مارکوفی است؟ چرا؟

یک مثال از زنجیره‌های مارکوف گسسته-زمان

- به‌طور واضح مدل وضع هوا یک DTMC است. چون:
  - رفتار آینده تنها به حالت جاری بستگی دارد.
  - زمان (روزهای هفته) و حالت (آفتابی، ابری و بارانی) هر دو گسسته هستند.
  - همچنین زمان نیز همگن است.
- DTMC وضع هوا دارای سه حالت است. اجازه دهید که ۱ را به آفتابی، ۲ را به ابری و ۳ را به بارانی منتسب کنیم. همچنین فرض کنید که زمان در جمعه صفر است. آنگاه:

$$\pi(0) = (0,0,1)$$

$$P = \begin{pmatrix} .6 & .3 & .1 \\ .4 & .45 & .15 \\ .15 & .6 & .25 \end{pmatrix}$$

حل مثال زنجیره‌های مارکوف گسسته-زمان

- وضع هوا در شنبه که با $\pi(1)$ نشان داده می‌شود عبارت است از:

$$\pi(1) = \pi(0)P = (0,0,1)\begin{pmatrix} .6 & .3 & .1 \\ .4 & .45 & .15 \\ .15 & .6 & .25 \end{pmatrix} = (.15,.6,.25),$$

که در نتیجه با احتمال ۰.۱۵ آفتابی، با احتمال ۰.۶ ابری و با احتمال ۰.۲۵ بارانی خواهد بود.

- وضع هوا در یکشنبه که با $\pi(2)$ نشان داده می‌شود عبارت است از:

$$\pi(2) = \pi(1)P = (.15,.6,.25)\begin{pmatrix} .6 & .3 & .1 \\ .4 & .45 & .15 \\ .15 & .6 & .25 \end{pmatrix} = (.3675,.465,.1675).$$

- وضع هوا در دوشنبه که با $\pi(3)$ نشان داده می‌شود عبارت است از:

$$\pi(3) = \pi(2)P = (.4316,\ .42,\ .1484),$$

که در نتیجه با احتمال ۰.۴۳ آفتابی، با احتمال ۰.۴۲ ابری و با احتمال ۰.۱۵ بارانی خواهد بود.

حل مثال زنجیره‌های مارکوف گسسته-زمان

- راه حل دیگر این است که P^3 را محاسبه کنیم. زیرا داریم:

$$\pi(3) = \pi(0)\, P^3$$

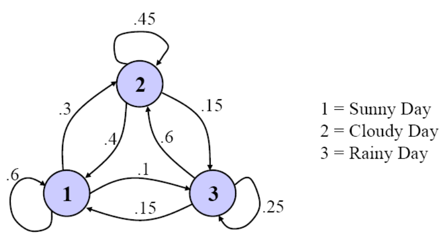
- حل دستی، دشوار و در معرض خطا است. مخصوصاً اگر مدل بزرگتر از مثال فوق باشد و دارای تعداد زیادی حالت باشد.
- از اینرو از بسته‌های نرم‌افزاری به‌طور گسترده برای این‌گونه تحلیل‌ها استفاده می‌شود (نظیر Maple یا MATLAB).
- این بسته‌های نرم‌افزاری $\pi(k)$ را بوسیله $(\ldots((\pi(0)P)P)P\ldots)P$ حساب می‌کنند، به‌جای آنکه P^k را حساب کنند. چرا؟

حل مثال زنجیره‌های مارکوف گسسته-زمان

- این بسته‌های نرم‌افزاری $\pi(k)$ را بوسیله $(\ldots((\pi(0)P)P)P\ldots)P$ حساب می‌کنند، به‌جای آنکه P^k را حساب کنند. چرا؟
  - محاسبه مستقیم P^k نیاز به حافظه بیشتری (به اندازه حداقل یک ماتریس) خواهد داشت.
  - در حالی که حاصل $\pi(0)P$ یک بردار ردیفی خواهد شد و $(\ldots((\pi(0)P)P)P\ldots)P$ نیز تماماً به یک بردار حافظه نیاز خواهد داشت.

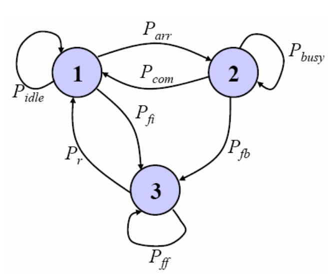
نمایش گرافیکی زنجیره‌های مارکوف

- اغلب در مورد زنجیره‌های مارکوف کوچک، مناسب است که DTMCها را با یک گراف جهت‌دار (directed graph) یا یک نمودار حالت-گذر (STD: state transition diagram) نشان دهیم:
  - رئوس متناظر با حالتها هستند.
  - یالها توسط احتمالات برچسب‌گذاری شده‌اند.
- برای مثال در مورد مدل پیش‌بینی وضع هوا داریم:



مثال: مدل مارکوف یک کامپیوتر ساده

- یک کامپیوتر یا بیکار (idle)، یا در حال کار (working) یا خراب (failed) است:
  - وقتي که کامپیوتر بیکار است، در یک سیکل ساعت (clock cycle)، برنامه‌ها با احتمال P_{arr} وارد می‌شوند.
  - وقتي که کامپیوتر در حال کار است، برنامه‌ها با احتمال P_{com} در یک سیکل ساعت کامل می‌شوند.
  - وقتي که کامپیوتر درحال کار است با احتمال P_{fb} و وقتی بیکار است با احتمال P_{fi} خراب می‌شود.
  - وقتي که کامپیوتر خراب است با احتمال P_r در یک سیکل ساعت تعمیر می‌شود.



$X = 1$	computer idle
$X = 2$	computer working
$X = 3$	computer failed

$$P = \begin{bmatrix} P_{idle} & P_{arr} & P_{fi} \\ P_{com} & P_{busy} & P_{fb} \\ P_r & 0 & P_{ff} \end{bmatrix}$$

رفتار حالت پایدار زنجیره‌های مارکوف گسسته-زمان

- ارزیابی کارآیی اغلب با رفتار سیستم‌ها پس از یک بازه زمانی طولانی سر و کار دارد.
- بازه زمانی طولانی با توجه به زمان بین رخدادها در سیستم تعریف می‌شود.
- از دیدگاه مدلسازی نکته فوق خیلی اهمیت دارد. زیرا مدلساز ممکن است که مجبور باشد که حالت اولیه سیستم را مشخص کند. با لحاظ نمودن رفتار سیستم در بلند مدت، تاثیر انتخاب دلخواه حالت اولیه در این حالت از بین رفته است.
- بنا بر این حالت پایدار (steady state) یا حالت تعادل (equilibrium) مدل یا سیستم متناظر با آن زمانی است که تاثیر حالت اولیه خنثی شده است.
- نکته فوق به این معنی نیست که سیستم وارد یک حالتی شده است که نمی‌تواند از آن گذر کند. بلکه به این معنی است که رفتار مدل دارای قانونمندی (regularity) یا قابلیت پیش‌بینی (predictability) در فضای حالت است.

رفتار حالت پایدار زنجیره‌های مارکوف گسسته-زمان

- اغلب مفید است که حد زمانی رفتار (time-limiting behavior) یک DTMC را بدانیم.
- منظور از "حد زمانی رفتار"، رفتار طولانی مدت (long-term) است، که سیستم به یک حالت پایدار (steady-state) رسیده است.
- به‌طور صوری ما به‌دنبال محاسبه $\lim_{n\to\infty} \pi(n)$ هستیم.
- در حقیقت برای محاسبه آن ما باید $\lim_{n\to\infty} \pi(0)P^n$ را محاسبه کنیم.
- که برای محاسبه آن راه‌حلهای مختلفی وجود دارد...

رفتار حالت پایدار زنجیره‌های مارکوف گسسته-زمان

- ساده‌ترین راه حل محاسبه حد، محاسبه $\pi(n)$ برای یک n خیلی بزرگ است.
- در آن صورت $\pi(n+1) \cong \pi(n)$ خواهد بود و می‌توانیم اعتقاد داشته باشیم که $\pi(n)$ تقریب خوبی از حالت پایدار است.
  - اما این روش در صورتی که n خیلی بزرگ باشد، ناکارآمد خواهد بود.
- اما داشتیم:

$$\pi(n+1) = \pi(n)\, P$$
$$\text{if } n \to \infty \quad \Rightarrow \pi(\infty+1) = \pi(\infty)\, P$$
$$\Rightarrow \pi(\infty) = \pi(\infty)\, P$$
$$\Rightarrow \pi^* = \pi^*\, P$$

- پس کافی است معادله خطی $\pi^* = \pi^* P$ یا $\pi=\pi P$ حل شود تا π^* یا π بدست آید که همان بردار احتمالات حالت پایدار است.
- راه حل اصلی پس از دسته‌بندیهای حالتهای زنجیره‌های مارکوف ارائه می‌شود.

دسته‌بندی حالتهای زنجیره‌های مارکوف گسسته-زمان

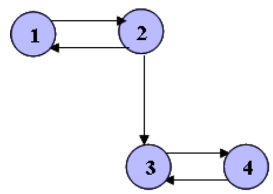
- حل حالت پایدار برخی از DTMCها از برخی دیگر بسیار آسانتر است.
- برای تعیین اینکه حل یک DTMC آسان است یا نه ما نیاز به ارائه تعدادی تعریف داریم:
- تعریف ۱: به یک حالت j، قابل دسترس (accessible or reachable) از حالت i گویند اگر یک $n\geq1$ وجود داشته باشد طوری که $P_{ij}^{(n)}>0$ باشد.
  - در آن صورت می‌نویسیم: $i \to j$
  - توجه: بیاد بیاورید که: $P_{ij}^{(n)} = P[X(n)=j|X(0)=i]$
  - دسترس‌پذیری در نمایش گرافیکی به این معنی است که یک مسیر از یالهای غیرصفر از گره i به گره j وجود داشته باشد. در آن صورت حالت j از حالت i قابل دسترس است.

دسته‌بندی حالتهای زنجیره‌های مارکوف گسسته-زمان

- **تعریف ۲:** به یک DTMC **کاهش‌ناپذیر** (irreducible) گویند اگر هر حالت آن از سایر حالتها قابل دسترس باشد.
- به‌طور صوری، یک DTMC کاهش‌ناپذیر است اگر:

$i \rightarrow j \quad \textit{for all } i, j \in S$

- یک DTMC **کاهش‌پذیر** (reducible) گویند اگر کاهش‌ناپذیر نباشد!
- مثال زیر یک DTMC کاهش‌پذیر است:

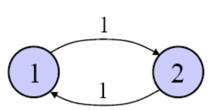


دسته‌بندی حالتهای زنجیره‌های مارکوف گسسته-زمان

- حل یک DTMC کاهش‌ناپذیر آسانتر است. برای این منظور لازم است که معادله خطی زیر حل شود:

$$\pi = \pi P$$

- دلیل درست بودن آنرا بعداً خواهیم دید.
- اما، قبل از آن مساله دیگری وجود دارد که باید مورد بحث قرار گیرد. DTMC زیر را در نظر بگیرید:



$\pi(0) = (1,0)$

□ آیا $\lim_{n \to \infty} \pi(n)$ وجود دارد؟ نه! چرا؟

دسته‌بندی حالتهای زنجیره‌های مارکوف گسسته-زمان

- تعریف ۳: به حالت i متناوب با دوره تناوب d گفته می‌شود اگر $P_{ii}^{(n)}>0$ فقط وقتی که n مضربی از d باشد.
- اگر d=1 باشد، گفته می‌شود که i نامتناوب (aperiodic) است:
  - یعنی اگر بزرگترین مقسوم‌علیه مشترک (GCD) دوره‌های تناوب مختلف برابر با یک باشد.
  - در مثال شکل فوق دوره‌های تناوب ۲، ۴، ۶، ... است که d=GCD(2, 4, 6, ...)=2 است. پس هر دو حالت متناوب هستند.
  - اما در مثال پیش‌بینی وضع هوا یا کامپیوتر ساده d = GCD(1,2,3, ..) = 1 است. پس همه حالتها نامتناوب هستند.
- حد رفتاری و در نتیجه حل حالت پایدار برای یک DTMC کاهش‌ناپذیر وجود دارد، اگر همه حالت‌های آن نامتناوب باشند.
- در خیلی از مراجع، به یک DTMC که کاهش‌ناپذیر، نامتناوب و بازگشتی غیرتهی (recurrent non-null) باشد، ارگودیک (ergodic) گفته می‌شود.

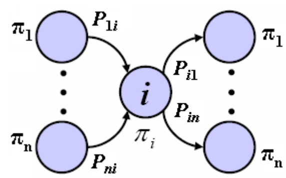
یعنی احتمال بازگشت به هر حالت در تعداد گام متناهی وجود داشته باشد

حل حالت پایدار DTMC

- رفتار حالت پایدار DTMC با حل معادله خطی $\pi = \pi P$ و با لحاظ نمودن این محدودیت که $\Sigma\pi_i = 1$ قابل محاسبه است:
  - برای DTMCهای کاهش‌ناپذیر می‌توان نشان داد که این حل یگانه (unique) است.
  - اگر یک DTMC کاهش‌ناپذیر است حاصل حل فوق π^* (احتمالات حالت پایدار) خواهد بود.
- اما نحوه بدست آوردن معادله خطی $\pi=\pi P$ با استفاده از معادلات جریان (flow equations) هم امکان پذیر است.
  - به معادلات جریان معادلات توازن سراسری (GBE: global balance equations) هم گفته می‌شود.
- این معادلات جریان نیاز به توضیح بیشتری دارند...

معادلات جریان

- گراف یک DTMC را تصور کنید که به گره‌ها احتمالات اشغال‌بودن (π_i) منتسب شده است:



  - شرط احتمال بودن π_iها باید برقرار باشد. یعنی: $\sum \pi_i = 1$

- اجازه دهید که $\pi_i P_{ij}$ "جرم احتمال" (probability mass) آن باشد که با یک گام از حالت i به حالت j برویم.
  - به جرم احتمالی، جریان (flow) یا شار (flux) هم گفته می‌شود.

معادلات جریان

- برای حالتی مثل i خواهیم داشت:

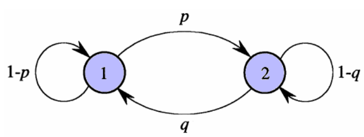
جریان خروجی از حالت i = جریان ورودی به حالت i

$$\sum_{j=1}^{n} \pi_j P_{ji} = \sum_{j=1}^{n} \pi_i P_{ij}$$
$$= \pi_i \sum_{j=1}^{n} P_{ij}$$
$$= \pi_i$$

- که نوشتن آن به شکل ماتریسی، همان $\pi P = \pi$ خواهد بود.

مثال: DTMC یک سیستم کامپیوتری با دو پردازنده و یک حافظه مشترک

- **احتمال دسترسی یکی از دو پردازنده به حافظه مشترک در هر سیکل ساعت (clock cycle) پردازنده برابر با p است. همچنین دسترسی به حافظه مشترک در یک سیکل ساعت با احتمال q خاتمه می‌یابد. آنگاه خواهیم داشت:**
  - **حالت ۱:** هیچکدام از دو پردازنده به حافظه مشترک دسترسی ندارند.
  - **حالت ۲:** یکی از دو پردازنده به حافظه مشترک دسترسی دارد.



  - آنگاه ماتریس P عبارت خواهد بود از:

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

مثال: DTMC یک سیستم کامپیوتری با دو پردازنده و یک حافظه مشترک

- **بردار احتمالات حالت پایدار:** $\pi = [\pi_1, \pi_2]$
- **آنگاه:** $\pi \mathbf{P} = \pi$

$$[\pi_1, \pi_2]\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} = [\pi_1, \pi_2]$$

$$\left| \begin{array}{l} \pi_1 (1-p) + \pi_2 q = \pi_1 \\ \pi_1 p + \pi_2 (1-q) = \pi_2 \\ \pi_1 + \pi_2 = 1 \end{array} \right.$$

- **پس از حل خواهیم داشت:**

$\pi_1 = q/(p+q)$ و $\pi_2 = p/(p+q)$

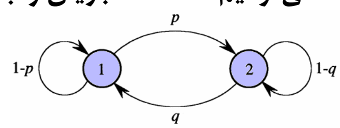
- **یا بردار احتمالات حالت پایدار (یا توزیع احتمالی حالت پایدار):**

$$\pi^* = [q/(p+q),\ p/(p+q)]$$

مثال: DTMC یک سیستم کامپیوتری با دو پردازنده و یک حافظه مشترک

- می‌توانیم معادلات جریان را به‌طور مستقیم هم بنویسیم:

"برای هر گره جریان وارد شونده با خارج شونده



$$\begin{cases}\pi_1 (1-p) + \pi_2 q = \pi_1 (1-p) + \pi_1 p \\ \pi_2 (1-q) + \pi_1 p = \pi_2 (1-q) + \pi_2 q \\ \pi_1 + \pi_2 = 1\end{cases} \Rightarrow \begin{cases}\pi_1 (1-p) + \pi_2 q = \pi_1 \\ \pi_1 q + \pi_2 (1-q) = \pi_2 \\ \pi_1 + \pi_2 = 1\end{cases}$$

- که همان معادلات قبلی است و پس از حل خواهیم داشت:

$\pi_1 = q/(p+q)$ و $\pi_2 = p/(p+q)$

- یا بردار احتمالات حالت پایدار (یا توزیع احتمالی حالت پایدار):

$$\pi^* = [q/(p+q),\ p/(p+q)]$$

- نوشتن معادلات جریان از روی زنجیره مارکوف روش آسانتری است.

زنجیره‌های مارکوف پیوسته-زمان

- در زنجیره‌های مارکوف گسسته-زمان، گذر از هر حالت به حالتی دیگر دارای زمان برابر است.
- اما در اغلب سیستم‌های مورد نظر، رخدادها در هر نقطه از زمان اتفاق می‌افتند. یعنی زمان گذر از هر حالت به حالتی دیگر طبق یک متغیر تصادفی پیوسته است.
- این امر باعث می‌شود که ما به مطالعه CTMCs بپردازیم.
- CTMC به‌صورت زیر تعریف می‌شود (خاصیت مارکوفی برای CTMC):

$$\begin{aligned}&P[X(t+\tau)=j \mid X(t)=i, X(t-t_1)=k_1, X(t-t_2)=k_2,\dots,X(t-t_n)=k_n]\\&\quad = P[X(t+\tau)=j \mid X(t)=i],\\&\quad = P_{ij}(\tau)\\&\text{for all } \tau>0,\ 0<t_1<t_2<\dots<t_n\end{aligned}$$

- مفهوم همگن بودن زمان در مورد CTMC هم وجود دارد:

$$P[X(t + \tau) = x_k \mid X(t) = x_j] = P[X(s + \tau) = x_k \mid X(s) = x_j]$$

خصوصیت‌های زنجیره‌های مارکوف پیوسته-زمان

- تعریف CTMC تا زمانی که خصوصیت‌های آنرا ندانیم خیلی مفید نخواهد بود.
- در ابتدا توجه کنید که $p_{ij}(\tau)$ مستقل از آن است که قبلاً چه مدت CTMC در حالت i بوده است. به این صورت که:

$$
\begin{aligned}
&P\left[X(t+\tau)=j \middle| X(u)=i \text{ for } u\in[0,t]\right] \\
&\quad = P\left[X(t+\tau)=j \middle| X(t)=i\right] \\
&\quad = p_{ij}(\tau)
\end{aligned}
$$

- تعریف فوق برای CTMC همان تعریف بی‌حافظه بودن (memoryless) است.
- تنها یک متغیر تصادفی پیوسته است که دارای خاصیت فوق است و آن متغیر تصادفی نمایی است.
- از اینجا مشخص می‌شود که CTMC با توزیع نمایی دارای رابطه است:
  - زمان گذر از حالتی به حالت دیگر دارای توزیع نمایی است.

ماتریس نرخ-گذر-حالت

- یک CTMC با استفاده از یک توزیع اولیه $\pi(0)$ و یک ماتریس گذر-نرخ (transition-rate matrix) که به آن ماتریس مولد بی‌نهایت کوچک (infinitesimal generator matrix) نیز گفته می‌شود، به‌طور کامل قابل توصیف است.
- ماتریس نرخ-گذر-حالت $Q = [q_{ij}]$ به‌صورت زیر تعریف می‌شود:

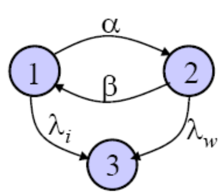
$$
q_{ij} = \begin{cases} \text{rate of going from state } i \text{ to state } j & i \neq j, \\ -\sum_{k\neq i} q_{ik} & i = j. \end{cases}
$$

یک مثال ساده از زنجیره‌های مارکوف پیوسته-زمان

- یک کامپیوتر یا بیکار (idle)، یا در حال کار (working) یا خراب (failed) است:
  - وقتی که کامپیوتر بیکار است، برنامه‌ها طبق یک توزیع نمایی با نرخ α وارد می‌شوند و طبق یک توزیع نمایی با نرخ β کامل می‌شوند.
  - وقتی که کامپیوتر درحال کار است طبق توزیع نمایی با نرخ λ_w و وقتی بیکار است طبق توزیع نمایی با نرخ λ_i خراب می‌شود.

یک مثال ساده از زنجیره‌های مارکوف پیوسته-زمان

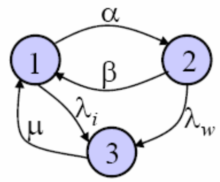
- حالتهای CTMC:
  - X=1 نشان‌دهنده حالت بیکار،
  - X=2 نشان‌دهنده حالت در حال کار و
  - X=3 نشان‌دهنده حالت خرابی.
- ماتریس Q:



$$Q = \begin{bmatrix} -(\alpha+\lambda_i) & \alpha & \lambda_i \\ \beta & -(\beta+\lambda_w) & \lambda_w \\ 0 & 0 & 0 \end{bmatrix}$$

یک مثال ساده از زنجیره‌های مارکوف پیوسته-زمان

- حال یک تغییر کوچک در مثال فوق می‌دهیم: اگر زمان تعمیر کامپیوتر فوق دارای توزیع نمایی با نرخ μ باشد.
- CTMC جدید به‌صورت زیر خواهد بود:



$$Q=\begin{bmatrix} -(\alpha+\lambda_i) & \alpha & \lambda_i \\ \beta & -(\beta+\lambda_w) & \lambda_w \\ \mu & 0 & -\mu \end{bmatrix}$$

یک مثال ساده از زنجیره‌های مارکوف پیوسته-زمان

- **سئوالات زیر در مورد مدل کامپیوتر ساده با استفاده از مدل فوق قابل پاسخ دادن است:**
  - □ قابلیت دسترسی (availability) کامپیوتر ساده در زمان t (*قابلیت دسترسی لحظه‌ای*) چقدر است؟
  - □ میانگین زمان تا خرابی (MTTF) کامپیوتر ساده چقدر است؟
  - □ زمان پاسخ کامپیوتر ساده چقدر است؟
  - □ توان عملیاتی (throughput) کامپیوتر ساده (تعداد کارهای پردازش‌شده در واحد زمان) با به‌حساب آوردن خرابیها و تعمیرات چقدر است؟
  - □ **و غیره.**

انواع حل زنجیره‌های مارکوف پیوسته-زمان

- سئوالاتی که وابسته به زمان هستند جزء معیارهای اتکاءپذیری بوده و با حل گذرای CTMC (transient solution) پاسخ داده می‌شوند.
  - نظیر: قابلیت دسترسی لحظه‌ای و MTTF.
- اما سئوالاتی که مربوط به حالت پایدار هستند جزء معیارهای کارآیی بوده و با حل حالت پایدار (steady-state solution) CTMC پاسخ داده می‌شوند.
  - نظیر: زمان پاسخ یا توان عملیاتی

حل گذاری زنجیره‌های مارکوف پیوسته-زمان

- رفتار گذاری CTMC به صورت زیر تعریف می‌شود:

$$\frac{d}{dt}\pi(t) = \pi(t)Q$$

- *این رابطه از کجا بدست آمده است؟*
- حل معادله دیفرانسیل فوق در برخی حالتها مشکل است.
- برای ارزیابی اتکاءپذیری و انجام‌پذیری، به حل معادله دیفرانسیل فوق و بدست آوردن توزیع π(t) نیازمند هستیم.

حل گذاری زنجیره‌های مارکوف پیوسته-زمان

- اما سه روش برای حل معادله دیفرانسیل فوق وجود دارد:
  1. با استفاده از فن تبدیل لاپلاس، معادله $\pi(t) = \pi(0)\, e^{Qt}$ به‌طور صریح حل شود؛
  2. با استفاده از روشهای انتگرال عددی، معادله $\pi'(t) = \pi(t)\, Q$ حل شود؛ یا
  3. رابطه بازگشتی $\pi(n+1) = \pi(n)\, Q$ حل شود.

- ما در این درس به حل گذرای CTMC نمی‌پردازیم و هدف حل حالت پایدار است که در مورد ارزیابی کارایی مورد نیاز است.
- مدل‌های کارآیی اغلب به حالت پایدار می‌رسند.

حل حالت پایدار زنجیره‌های مارکوف پیوسته-زمان

- مفهوم کاهش‌ناپذیری در مورد CTMC هم مطرح است. اما مفهوم نامتناوب بودن معنی ندارد. زیرا همه گره‌های CTMC دارای یک کمان به طول به خود هستند و مدتی را که به آن زمان اقامت (residence time or sojourn time) گفته می‌شود، در هر حالت می‌مانند. این زمان اقامت دارای توزیع نمایی منفی است. از اینرو، در نمودار گذر-حالت CTMC، چنین کمانهایی اصلاً نشان داده نمی‌شوند.
  - بنا بر این، برای ارگودیک بودن یک CTMC، شرط کاهش‌ناپذیر کفایت می‌کند.
- چنانچه یک CTMC ارگودیک باشد، حل حالت پایدار CTMC به‌صورت زیر می‌تواند انجام شود: $\pi^* Q=0$
  - که در آن: $\pi^* = \lim_{t\to\infty} \pi(t)$

حل حالت پایدار زنجیره‌های مارکوف پیوسته-زمان

- یک راه رسیدن به $\pi^* Q=0$، استفاده از معادلات جریان است:

جریان جرم احتمالی وارد شونده به حالت i = جریان جرم احتمالی خارج‌شونده از حالت i

  - جریان جرم احتمالی خارج‌شونده از حالت i، به‌طور ساده برابر است با $\pi_i q_{ij}$، که احتمال بودن در حالت i ضرب در نرخ گذر (انتقال) از حالت i به حالت j است.

Flow into state i: $\sum_{\substack{j=1\\j\neq i}}^{n} \pi_j q_{ji}$

Flow out of state i: $\sum_{\substack{j=1\\j\neq i}}^{n} \pi_i q_{ij} = \pi_i \sum_{\substack{j=1\\j\neq i}}^{n} q_{ij} = \pi_i(-q_{ii})$

$$\longrightarrow \quad \sum_{\substack{j=1\\j\neq i}}^{n} \pi_j q_{ji} = \pi_i(-q_{ii})$$

$$\sum_{\substack{j=1\\j\neq i}}^{n} \pi_j q_{ji} + \pi_i q_{ii} = 0$$

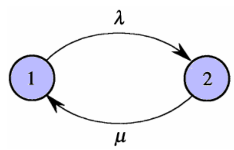
$$\sum_{j=1}^{n} \pi_j q_{ji} = 0$$

حل حالت پایدار زنجیره‌های مارکوف پیوسته-زمان

- وقتی که CTMC کاهش‌ناپذیر باشد، آنگاه با $\sum_{i=1}^{n} \pi_i^* = 1$ محدودیت می‌توان π^* را محاسبه نمود.
- اما وقتی که CTMC کاهش‌ناپذیر نباشد، روشهای حل پیچیده‌ای مورد نیاز است.
- دستگاه معادلات خطی حالت پایدار را به‌صورت $\pi^* Q = 0$ می‌نویسیم.
- حل دستگاه فوق به یک نتیجه واحد منجر نمی‌شود (دستگاه معادلات دارای چندین جواب است).
- از اینرو محدودیت فوق را اصافه می‌کنیم تا داشتن جواب یکتا را تضمین کنیم.

مثال: CTMC یک سیستم کامپیوتری دارای دو پردازنده و یک حافظه مشترک

- **توزیع زمان بین صدور درخواستهای دسترسی پردازنده‌ها به حافظه مشترک نمایی با نرخ λ بوده و زمان دسترسی به حافظه مشترک دارای توزیع نمایی با نرخ μ است. آنگاه:**
  - **حالت ۱:** هیچکدام از دو پردازنده به حافظه مشترک دسترسی ندارند.
  - **حالت ۲:** یکی از دو پردازنده به حافظه مشترک دسترسی دارد.



$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix}$$

مثال: حل CTMC یک سیستم کامپیوتری دارای دو پردازنده و یک حافظه مشترک

- **حال دستگاه معادلات جریان را نوشته و حل می‌کنیم:**

$$\begin{cases} 0 = -\lambda\pi1 + \mu\,\pi2 \\ 0 = \lambda\pi1 - \mu\,\pi2 \\ 1 = \pi1 + \pi2 \end{cases}$$

$$\Rightarrow \pi1 = \mu/(\lambda+\mu) \qquad \pi2 = \lambda/(\lambda+\mu)$$

$$\Rightarrow \pi^* = [\mu/(\lambda+\mu),\ \lambda/(\lambda+\mu)]$$

- ***معیارهای قابل محاسبه:***
  - بهره‌وری حافظه مشترک؟
  - توان عملیاتی حافظه مشترک؟
  - میانگین تعداد پردازنده‌های در حال دسترسی به حافظه مشترک؟
  - میانگین تعداد پردازنده‌های منتظر حافظه مشترک؟
  - میانگین زمان پاسخ حافظه مشترک؟
  - میانگین زمان انتظار حافظه مشترک؟

مثال: حل CTMC یک سیستم کامپیوتری دارای دو پردازنده و یک حافظه مشترک

- *معیارهای قابل محاسبه:*
  - بهره وری حافظه مشترک؟ $\pi_2 = \lambda/(\lambda+\mu)$
  - توان عملیاتی حافظه مشترک؟ $\pi_2\mu = \lambda\mu/(\lambda+\mu)$
  - میانگین تعداد پردازنده های در حال دسترسی به حافظه مشترک؟ π_2
  - میانگین تعداد پردازنده‌های منتظر حافظه مشترک؟ 0
  - میانگین زمان پاسخ حافظه مشترک؟ $1/\mu$
  - میانگین زمان انتظار حافظه مشترک؟ 0

حل عددی زنجیره‌های مارکوف پیوسته-زمان

- انواع روشهای حل عددی دستگاه معادلات که در نرم‌افزارهای جبر خطی استفاده می‌شوند عبارتند از:
  - روشهای مستقیم (direct methods)
  - روشهای تکراری (iterative methods)

روشهای مستقیم حل دستگاه معادلات

- راه حل مستقیم حل دستگاه معادلات خطی، استفاده از روش حذفی گاوس (Gauss-elimination) یا (Gauss–Jordan elimination) است.
- مزایای این روشها عبارتند از:
  - روشهای عددی مناسبی وجود دارد که بدست آوردن جواب را تضمین می‌کنند.
  - کارآیی روش حل قابل پیش‌بینی است.
  - بسته‌های نرم‌افزاری متعددی برای این روشها وجود دارد.
- عیب روش فوق این است که اغلب ماتریس Q، یک ماتریس بسیار بزرگ است که دارای هزاران و بلکه میلیونها حالت بوده و اغلب خلوت (sparse) است. در نتیجه روش ضمن نیاز به حافظه فراوان دارای کارایی ضعیفی است.

روشهای تکراری حل دستگاه معادلات

- روشهای تکراری ایستا مثل روش ژاکوبی (Jacobbi) و گاوس-سایدل (Gauss-Seidel)، روشهایی هستند که می‌توان آنها را به‌صورت زیر نوشت:

$$\pi^{(k+1)} = \pi^{(k)} M$$

که در آن π بردار جواب و M یک ماتریس ثابت (ایستا) است. آنگاه جواب دستگاه معادلات با π^* نشان داده می‌شود.

- محاسبه $\pi^{(k+1)}$ از $\pi^{(k)}$ به یک ضرب بردار در ماتریس نیاز دارد، که یک تکرار (iteration) نامیده می‌شود.

روشهای تکراری ایستای حل دستگاه معادلات

- در مورد DTMC داشتیم: $\pi^* = \pi^* P$
  - در این صورت $M = P$ و می‌توانیم داشته باشیم: $\pi^{(k+1)} = \pi^{(k)} P$
- اما در مورد CTMC داریم: $\pi^* Q = 0$
  - حال می‌توانیم فرض کنیم: $M = Q + I$ که در آن I یک ماتریس قطری یکّه است.
  - در این صورت خواهیم داشت:

$$\pi^* Q = 0 \quad => \quad \pi^* (M-I) = 0 \quad => \pi^* M = \pi^*$$

  - در نتیجه می‌توانیم داشته باشیم: $\pi^{(k+1)} = \pi^{(k)} M$

روش تکراری گاوس-سایدل

- برای دستگاه معادلات $A\ x = b$ روش گاوس-سایدل:

$$\mathbf{x}^{(k)} = (D - L)^{-1} (U\,\mathbf{x}^{(k-1)} + \mathbf{b}),$$

- که در آن:
  - D ماتریس قطری (diagonal) ماتریس A است.

$$D = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix},$$

  - L ماتریس مثلثی کران پایین محکم (strictly lower triangular):

$$L = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ a_{n1} & a_{n2} & \cdots & 0 \end{bmatrix}.$$

  - و U ماتریس مثلثی کران بالای محکم (strictly upper triangular):

$$U = \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

روش تکراری گاوس-سایدل

- برای این منظور خواهیم داشت : $\pi^{(k+1)} = \pi^{(k)} U (D+L)^{-1}$
- آنگاه الگوریتم زیر برای حل قابل استفاده است:

for $k = 1$ to convergence

 for $i = 1$ to n

$$\pi_i^{(k+1)} = -\frac{1}{q_{ii}}\left(\sum_{j=1}^{i-1} \pi_j^{(k+1)} q_{ji} + \sum_{j=i+1}^{n} \pi_j^{(k)} q_{ji} \right)$$

 end for

end for

- خواص روش گاوس-سایدل
  - از نظر حافظه کارآمد است. زیرا فقط به Q و π نیاز دارد.
  - سریعتر از روش ژاکوبی همگرا می‌شود.

جمع‌بندی

- در این فصل ابتدا مفاهیم و تعاریف اولیه فرآیندهای مارکوف را ارائه نمودیم.
- آنگاه زنجیره‌های مارکوف گسسته-زمان را معرفی نموده و روشهای حل گذرا و حالت پایدار آنها را ارائه کردیم.
- سپس زنجیره‌های مارکوف پیوسته-زمان و روشهای حل گذرا و حالت پایدار آنها را معرفی نمودیم.
- در خاتمه هم اشاره‌ای به روشهای حل عددی زنجیره‌های مارکوف نمودیم.