دانشگاه تهران، دانشکده مهندسی برق و کامپیوتر

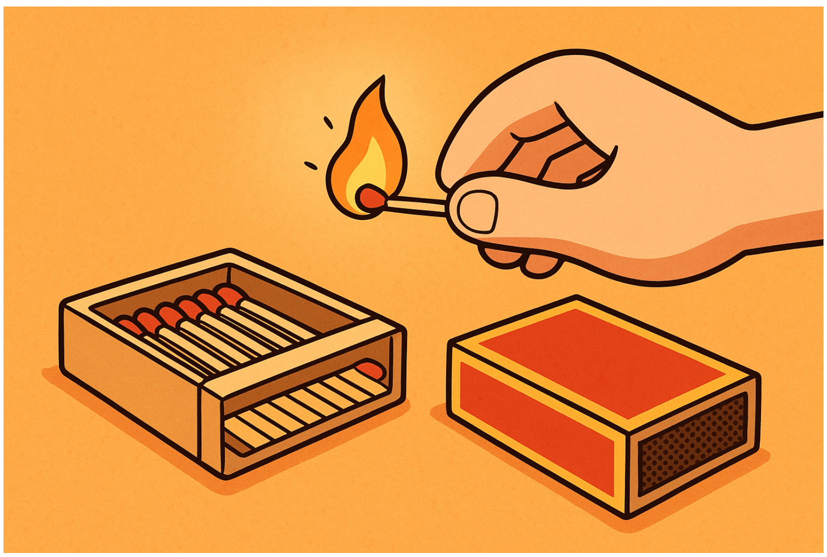
آمار و احتمال مهندسی

تمرین کامپیوتری سوم - توزیع های گسسته

طراح: مریم ولی

۱. مسئله کبریت باناخ

فرض کنید دو جعبه کبریت داریم که هر یک در ابتدا دارای n عدد کبریت هستند. در هر مرحله، یکی از دو جعبه با احتمال برابر $\frac{1}{2}$ انتخاب می‌شود و یک کبریت از آن برداشته می‌شود. این فرایند تا زمانی ادامه می‌یابد که یکی از جعبه‌ها برای نخستین بار خالی شود.



می‌خواهیم بدانیم در لحظه‌ای که یکی از جعبه‌ها خالی می‌شود، احتمال آن‌که جعبه‌ی دیگر هنوز دقیقاً k عدد کبریت داشته باشد چقدر است.

به‌عبارت دیگر، اگر متغیر تصادفی K تعداد کبریت‌های باقی‌مانده در جعبه‌ی غیرخالی را در زمان خالی شدن جعبه‌ی دیگر نشان دهد، هدف ما یافتن توزیع احتمال $P(K = k)$ است.

ساختار تصادفی مسئله

در هر گام از فرایند، انتخاب بین دو جعبه یک متغیر تصادفی از نوع **برنولی** است. اگر انتخاب جعبه‌ی اول را با مقدار ۱ و انتخاب جعبه‌ی دوم را با مقدار ۰ نمایش دهیم، آنگاه داریم:

$$X_t \sim \text{Bernoulli}\left(p = \tfrac{1}{2}\right),$$

که در آن X_t تصمیم انتخاب در گام t را نشان می‌دهد.

پس از انتخاب جعبه، از میان کبریت‌های موجود در آن جعبه یکی به‌طور یکنواخت تصادفی برداشته می‌شود. بنابراین انتخاب کبریت‌ها دارای توزیع **یکنواخت گسسته** در مجموعه‌ی کبریت‌های آن جعبه است. اما از آن‌جا که تنها تعداد کبریت‌ها برای ما اهمیت دارد (و نه هویت آن‌ها)، این انتخاب یکنواخت تأثیری در توزیع گذار بین حالت‌ها ندارد و تنها کاهش یک واحد در شمار کبریت‌های همان جعبه را رقم می‌زند.

ویژگی مارکوفی (Markov Property)

در هر لحظه، وضعیت سیستم با جفت مرتب زیر مشخص می‌شود:

$$S_t = (x_۱, x_۲),$$

که در آن $x_۱$ و $x_۲$ به‌ترتیب تعداد کبریت‌های باقی‌مانده در جعبه‌ی اول و دوم هستند.

انتقال بین حالت‌ها به‌صورت زیر انجام می‌شود:

$$S_{t+۱} = \begin{cases} (x_۱ - ۱,\, x_۲) & \text{با احتمال } \frac{۱}{۲}, \\ (x_۱,\, x_۲ - ۱) & \text{با احتمال } \frac{۱}{۲}. \end{cases}$$

از آن‌جا که احتمال گذار به حالت بعدی تنها به حالت فعلی بستگی دارد و نه به کل تاریخچه‌ی گذشته، این فرایند یک **زنجیره‌ی مارکوف گسسته در زمان** محسوب می‌شود.

هدف

هدف نهایی، محاسبه‌ی توزیع احتمال $P(K = k)$ است؛ یعنی احتمال آن‌که در زمان خالی شدن یکی از جعبه‌ها، جعبه‌ی دیگر دارای k کبریت باشد.

۲. شبیه سازی صف در کافه

فاز مقدماتی: مرور کلی پروژه

این پروژه به توسعه‌ی یک مدل گسسته‌ی احتمالاتی و شبیه‌سازی برای یک سیستم کافه می‌پردازد که در آن مشتریان به‌صورت تصادفی وارد می‌شوند، از میان یک منوی شامل **۷ آیتم** سفارش می‌دهند، و منتظر آماده‌سازی سفارش خود می‌مانند.

هر مشتری:

۱. مطابق با توزیع **پواسون** با پارامتر نرخ λ وارد می‌شود؛

۲. بر اساس توزیع **چندجمله‌ای (چندوجهی)** ترجیح، یکی از آیتم‌های منو را انتخاب می‌کند؛

۳. زمان آماده‌سازی (سرویس) او از یک توزیع **هندسی** تبعیت می‌کند که به سختی آیتم انتخاب‌شده بستگی دارد.

کافه دارای دو سرویس دهنده اصلی است و زمانی که طول صف از آستانه‌ی T فراتر رود، سرویس دهنده سوم فعال می‌شود. در این پروژه، رفتار صف، دینامیک سرویس، احتمال‌های بلندمدت و حساسیت به پارامترها بررسی می‌شود.

فاز ۱: مقادیر نظری و شبیه‌سازی‌های کوچک

۱. محاسبه‌ی تابع جرم احتمال (PMF) و تابع توزیع تجمعی (CDF) توزیع پواسون با $\lambda = ۰/۸$ برای $k = ۰:۱۲$؛

۲. نمایش تابع احتمال دسته‌ای (categorical PMF) مربوط به آیتم‌های منو (q_i)؛

۳. نمایش نمونه‌ای از PMF و CDF توزیع هندسی برای یک مقدار دلخواه p با بازه‌ی ۱ تا ۲۰؛

۴. مقایسه‌ی هیستوگرام داده‌های شبیه‌سازی‌شده از پواسون با PMF نظری آن.

فاز ۲: نمونه‌گیری گسترده و تجسم (با درنظرگرفتن منو)

۱. تولید $N = ۲۰۰۰$ نمونه‌ی ورود پواسونی با λ پایه و محاسبه‌ی میانگین و واریانس تجربی؛

۲. تولید $N = ۲۰۰۰$ نمونه از منوی دسته‌ای و محاسبه‌ی نسبت‌های تجربی انتخاب آیتم‌ها؛

۳. برای هر آیتم منو، تولید ۱۰۰۰ نمونه‌ی هندسی (زمان سرویس) و ارائه‌ی:

- هیستوگرام‌های جداگانه (به‌صورت facet)،
- نمودارهای boxplot برای هر آیتم،
- نمودارهای ECDF برای سه آیتم نمونه.

۴. نمایش PMF توزیع پواسون برای طول بازه‌های زمانی ۵، ۱۰ و ۳۰ دقیقه با $\lambda_t = \lambda \cdot t/۵$.

فاز ۳: مقادیر تحلیلی هر آیتم و شناسایی گلوگاه‌ها

۱. محاسبه‌ی امید ریاضی و واریانس زمان سرویس برای هر آیتم منو با استفاده از روابط:

$$E[T_i] = \frac{۱}{p_i}, \qquad \mathrm{Var}(T_i) = \frac{۱ - p_i}{p_i^۲}.$$

۲. ترسیم نمودار میله‌ای زمان سرویس مورد انتظار برای هر آیتم؛

۳. شناسایی آیتم‌های گلوگاهی بر اساس مقدار $E[T_i]$ و $\mathrm{Var}(T_i)$؛

۴. : توضیح دهید که توزیع هندسی را می‌توان به‌عنوان یک حالت خاص از توزیع **دوجمله‌ای منفی** در نظر گرفت، در این بیان، T_i همان تعداد تلاش‌ها تا نخستین موفقیت است.

فاز ۴: شبیه‌سازی جزئی صف به‌صورت FIFO با جزئیات آیتمی

در این فاز مدل صف گسسته با جزئیات مربوط به آیتم‌ها گسترش می‌یابد و رفتار سیستم تحت نرخ‌های مختلف ورود بررسی می‌شود. ورود مشتریان از فرآیند پواسون(λ) در هر بازه‌ی زمانی تبعیت می‌کند و هر مشتری یکی از آیتم‌ها را با احتمال‌های q_i انتخاب می‌کند.

هر سفارش به تعداد تصادفی $T_i \sim \mathrm{Geom}(p_i)$ بازه زمان نیاز دارد. در هر بازه:

- دو سرویس دهنده پایه فعال هستند؛
- خدمت‌دهنده‌ی سوم زمانی فعال می‌شود که طول صف پیش از سرویس از حد h فراتر رود.

سپس سه سناریو با بار کم، متوسط و زیاد بررسی می‌شوند و خروجی‌های زیر تحلیل می‌گردند:

۱. نمودار سری زمانی طول صف (Q_t)؛

۲. توزیع تجربی احتمال صف (PMF تجربی Q)؛

۳. نمودار ناحیه‌ای ترکیبی از ترکیب آیتم‌ها در صف (۴۰۰ بازه‌ی اول)؛

۴. شاخص‌های خلاصه شامل:

- $\mathbb{E}[Q]$
- $\mathrm{Var}(Q)$
- $P(Q = ۰)$
- P(فعال بودن سرویس دهنده سوم).

فاز ۵: تحلیل حساسیت و پایداری نسبت به نرخ ورود λ

در این فاز، مطالعه‌ی سیستماتیک حساسیت نسبت به نرخ ورود λ انجام می‌شود تا مرزهای پایداری و عملکرد سیستم مشخص گردد. با استفاده از شبیه‌سازی فاز ۴، مقدار λ تا حدود ۵.۱ افزایش داده می‌شود و رفتار آماری صف اندازه‌گیری می‌گردد.

اهداف:

۱. تعیین ناحیه‌ی پایداری؛

۲. بررسی تغییرات $\mathbb{E}[Q]$، $\mathrm{Var}(Q)$، $P(Q = ۰)$ و P(فعال بودن خدمت‌دهنده‌ی سوم) با تغییر λ؛

۳. شناسایی نقطه‌ی بحرانی λ_{stab} که در آن پایداری از بین می‌رود.

معیار پایداری:

$$\mathbb{E}[Q] < ۰/۸ \times s_{\max} = ۲/۴, \quad s_{\max} = ۳.$$

فاز ۶: حساسیت نسبت به آستانه‌ی فعال‌سازی پویا h

در این فاز اثر پارامتر h بر پایداری صف بررسی می‌شود. خدمت‌دهنده‌ی سوم هنگامی فعال می‌شود که $Q_{t-۱} - ۲ \geq h$. h کوچک موجب پایداری بیشتر ولی مصرف منابع بالاتر می‌شود، در حالی‌که h بزرگ صرفه‌جویی در منابع را به قیمت ازدحام بیشتر به همراه دارد.

با ثابت نگه داشتن $\lambda = ۰/۸$ و طول شبیه‌سازی ۲۰۰۰ بازه، مقادیر زیر برای $h = ۱$ تا ۸ محاسبه می‌شوند:

۱. $E[Q]$، $Var(Q)$، $P(Q=۰)$، P(فعال بودن سوم)،

۲. و شاخص پایداری (پایدار اگر $E[Q] < ۲/۴$).

فاز ۷: تحلیل حساسیت مشترک نسبت به λ و h

در فاز نهایی، اثر ترکیبی نرخ ورود λ و آستانه‌ی فعال‌سازی h بر پایداری سیستم بررسی می‌شود. هدف، ترسیم نقشه‌ی پایداری دوبعدی است که نواحی رفتار پایدار صف را نشان دهد.

پارامترها:

$$\lambda \in \{۰/۲, ۰/۴, ۰/۶, ۰/۸, ۱/۰, ۱/۲, ۱/۴\}, \quad h \in \{۱, ۲, ۳, ۴, ۵, ۶, ۷, ۸\}.$$

برای هر ترکیب، شبیه‌سازی ۱۰۰۰ بازه انجام شده و مقادیر $E[Q]$، $Var(Q)$ و شاخص پایداری محاسبه می‌شود.

موفق باشید.