دانشکده علوم ریاضی
دانشگاه صنعتی شریف

الگوریتم‌های خلاصه‌سازی برای مه‌داده

محمدهادی فروغمنداعرابی
پاییز ۱۳۹۹

برخی کران‌های احتمالاتی

جلسه‌ی دوم

نگارنده: علیرضا توفیقی محمدی

۱ مروری بر مباحث گذشته

در جلسه‌ی گذشته، به معرفی درس پرداختیم؛ مباحثی که در درس مطرح خواهند شد، منابع درس و چگونگی ارائه‌ی درس را آموختیم.

در ادامه برخی از کران‌های احتمالاتی که در این درس مورد استفاده قرار خواهند گرفت را معرفی کنیم و مثالی از آن‌ها خواهیم دید.

۲ مروری بر نشانه‌گذاری‌های احتمال

ما به طور عمده با متغیرهای تصادفی گسسته کار می‌کنیم. یک متغیر تصادفی را با یک حرف بزرگ از الفبای انگلیسی مثل X نشان می‌دهیم. همچنین فرض می‌کنیم X مقادیرش را از مجموعه‌ای متناهی چون S می‌گیرد.

تعریف ۱ (امیدریاضی یک متغیر تصادفی). به یادآورید که امیدریاضی متغیر تصادفی X به شکل زیر تعریف می‌شود:

$$\mathbb{E}X = \sum_{j \in S} j \cdot \mathbb{P}r(X = j) \tag{۱}$$

تعریف ۲ (واریانس یک متغیر تصادفی). واریانس متغیر تصادفی X به شکل زیر تعریف می‌شود:

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \mathbb{E}[X^2] - \mathbb{E}X^2 \tag{۲}$$

تا اینجا با تعریف متغیرتصادفی، امیدریاضی و واریانس آن آشنا شدیم، در ادامه یکی از مهمترین ویژگی‌های امیدریاضی را خواهیم دید.

لم ۳ (خطی‌بودن امید ریاضی). برای هر دو عدد حقیقی a و b و هر دو متغیر تصادفی X و Y داریم:

$$\mathbb{E}[aX + bY] = a \cdot \mathbb{E}X + b \cdot \mathbb{E}Y \tag{۳}$$

خاصیت خطی‌بودن امیدریاضی این اجازه را به ما می‌دهد که امیدریاضی را روی جمع پخش کنیم، اما این‌کار برای ضرب در حالت کلی ممکن نیست، تعریف زیر به این منظور ارائه می‌شود.

تعریف ۴ (استقلال متغیرهای تصادفی). گوییم دو متغیر تصادفی X و Y با مقادیر S_X و S_Y مستقل‌اند، هرگاه:

$$\mathbb{P}r(X=x, Y=y) = \mathbb{P}r(X=x) \cdot \mathbb{P}r(Y=y) \qquad (۴)$$

حال خواهیم دید که با شرط استقلال، امیدریاضی روی ضرب نیز پخش می‌شود.

لم ۵ (امید ریاضی ضرب دو متغیر تصادفی مستقل). اگر X و Y دو متغیر تصادفی مستقل باشند، خواهیم داشت:

$$\mathbb{E}[X \cdot Y] = \mathbb{E}X \cdot \mathbb{E}Y \qquad (۵)$$

اثبات. برای اثبات کافی‌ست از طرف چپ شروع کنیم و به طرف راست تصادفی برسیم و از تعریف امیدریاضی و استقلال دو متغیر تصادفی استفاده کنیم.

$$\begin{aligned}
\mathbb{E}[X \cdot Y] &= \sum_{x,y} \mathbb{P}r(X=x, Y=y) \cdot x \cdot y \\
&= \left(\sum_x x \cdot \mathbb{P}r(X=x)\right) \left(\sum_y y \cdot \mathbb{P}r(Y=y)\right) \\
&= \mathbb{E}X \cdot \mathbb{E}Y
\end{aligned}$$

□

همانند امیدریاضی ضرب متغیرهای تصادفی، واریانس جمع متغیرهای تصادفی نیز در حالت کلی پخش نخواهد شد، اما لم زیر نشان می‌دهد با شرط استقلال، واریانس روی جمع پخش خواهد شد.

لم ۶ (واریانس جمع دو متغیر تصادفی مستقل). اگر X و Y دو متغیر تصادفی مستقل باشند، خواهیم داشت:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \qquad (۶)$$

اثبات.

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}\left[(X+Y)^۲\right] - \mathbb{E}[X+Y]^۲ \\
&= \mathbb{E}[X^۲] + \mathbb{E}[Y^۲] + ۲ \cdot \mathbb{E}[XY] - \mathbb{E}[X]^۲ - \mathbb{E}[Y]^۲ - ۲\mathbb{E}X \cdot \mathbb{E}Y \qquad (\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]) \\
&= \mathbb{E}[X^۲] - \mathbb{E}[X]^۲ + \mathbb{E}[Y^۲] - \mathbb{E}[Y]^۲ \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

□

۳ معرفی سه کران احتمالاتی

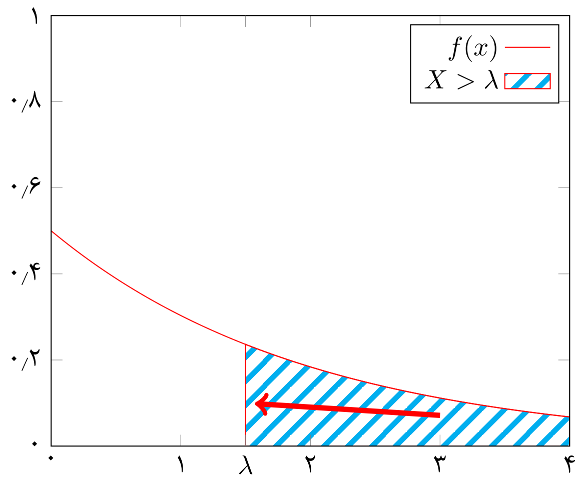
در ادامه به معرفی سه کران احتمالاتی روی متغیرهای تصادفی می‌پردازیم و اثباتی از آن‌ها ارائه می‌کنیم.

۱.۳ کران مارکوف

کران مارکوف اولین کران کاربردی در این درس است، با کمک آن، می‌توانیم احتمال یک متغیر تصادفی را با کمک امیدریاضی آن محدود کنیم.

لم ۷ (مارکوف). اگر X یک متغیر تصادفی با مقادیر نامنفی باشد، آن‌گاه برای هر $\lambda > ۰$ خواهیم داشت:

$$\mathbb{P}r(X > \lambda) < \frac{\mathbb{E}X}{\lambda} \qquad (۷)$$



شکل ۱: نمودار توزیع نمایی با پارامتر ۰/۵، مساحت هاشورخورده $\mathbb{P}r(X > \lambda)$ است، اگر کل این احتمال را بر روی λ جمع‌کنیم، امیدریاضی متغیر تصادفی به حداقل $\lambda \cdot \mathbb{P}r(X > \lambda)$ کاهش خواهد یافت.

اثبات. همان‌طور که در شکل ۱ می‌بینید، اگر به نمودار توزیع متغیر تصادفی X نگاه کنیم، این متغیر تصادفی مقادیری را می‌پذیرد که برخی از آن‌ها بیش از λ و برخی کم‌تر از λ است، امیدریاضی X برابر با مجموع حاصل ضرب مقادیر X در احتمال آن‌هاست، پس اگر تمام جرم توزیع که بیش از λ است را روی λ جمع کنیم، مقدار امیدریاضی لاآقل $\mathbb{P}r(X > \lambda) \cdot \lambda$ خواهد شد و درنتیجه

$$\mathbb{E}X > \mathbb{P}r(X > \lambda) \cdot \lambda$$

یا به طور دقیق‌تر:

$$\begin{aligned}
\mathbb{E}X &= \sum_x \mathbb{P}r(X = x) \cdot x \\
&= \sum_{x \leq \lambda} \mathbb{P}r(X = x) \cdot x + \sum_{x > \lambda} \mathbb{P}r(X = x) \cdot x \\
&> \sum_{x > \lambda} \mathbb{P}r(X = x) \cdot \lambda \\
&= \lambda \sum_{x > \lambda} \mathbb{P}r(X = x) \\
&= \lambda \mathbb{P}r(X > \lambda)
\end{aligned}$$

و نامساوی مارکوف ثابت شد. □

۲.۳ کران چبیشف

دومین کران مورد مطالعه‌ی ما، کران چبیشف است، در این نامساوی احتمال فاصله‌داشتن از امیدریاضی را محدود می‌کنیم و ثابت می‌کنیم احتمال فاصله‌ی زیاد از میانگین برای متغیرهای تصادفی کم است.

لم ۸ (چبیشف). اگر X یک متغیر تصادفی و $\lambda > 0$ یک عدد مثبت باشد، داریم:

$$\mathbb{P}r\left(|X-\mathbb{E}X| > \lambda\right) \leq \frac{\mathbb{E}\left[(X-\mathbb{E}X)^2\right]}{\lambda^2} = \frac{\text{Var}(X)}{\lambda^2} \tag{8}$$

اثبات. چون X یک متغیر تصادفی نامنفی است، داریم:

$$|X-\mathbb{E}X| > \lambda \iff (X-\mathbb{E}X)^2 > \lambda^2$$

پس با توجه به مشاهده‌ی قبل از نامساوی مارکوف استفاده می‌کنیم، متغیر تصادفی $Y = (X-\mathbb{E}X)^2$ را در نظر بگیرید. واضح است که Y یک متغیر تصادفی نامنفی است، پس در نامساوی مارکوف صدق می‌کند. همچنین:

$$|X-\mathbb{E}X| > \lambda \iff (X-\mathbb{E}X)^2 > \lambda^2 \iff Y > \lambda^2$$

پس:

$$\begin{aligned}
\mathbb{P}r\left(|X-\mathbb{E}X| > \lambda\right) &= \mathbb{P}r(Y > \lambda^2) \\
&< \frac{\mathbb{E}Y}{\lambda^2} && \text{(کران مارکوف)} \\
&= \frac{\mathbb{E}[(X-\mathbb{E}X)^2]}{\lambda^2} \\
&= \frac{\text{Var}(X)}{\lambda^2}
\end{aligned}$$

و نابرابری ثابت شد. □

۳.۳ کران چرنوف

تا اینجا کران‌های مارکوف و چبیشف را دیدیم که کرانی را براساس یک متغیر تصادفی به ما ارائه می‌دادند، در ادامه کران چرنوف را معرفی می‌کنیم که کرانی روی احتمال فاصله از میانگین مجموع تعدادی از متغیرهای تصادفی است.

لم ۹ (چرنوف). فرض کنید $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی مستقلی هستند که $X_i \in [0,1]$. $X = \sum_{i=1}^n X_i$ بگیرید که $\mu = \mathbb{E}X$ است. اگر $0 < \epsilon < 1$ باشد، آنگاه:

$$\mathbb{P}r\left(|X-\mathbb{E}X| > \epsilon\mu\right) \leq 2 \cdot e^{-\frac{\epsilon^2\mu}{3}} \tag{9}$$

اثبات. از اثبات نامساوی به صورت کلی صرف نظر می‌کنیم و تنها برای زمانی‌که هر X_i متغیر تصادفی برنولی با پارامتر p_i (با احتمال p_i یک و با احتمال $1-p_i$ صفر را بپذیرد) اثباتی ارائه می‌دهیم. همچنین اثبات آن را به اثبات دو نامساوی تقسیم کرده و تنها یکی از آن‌ها را، آن‌هم با ضریب ثابتی غیر از $\frac{1}{3}$، نشان می‌دهیم. داریم:

$$\begin{aligned}
\mathbb{P}r\left(|X-\mathbb{E}X| > \epsilon\mu\right) &= \mathbb{P}r\left(X-\mathbb{E}X > \epsilon\mu\right) + \mathbb{P}r\left(X-\mathbb{E}X < -\epsilon\mu\right) \\
&= \mathbb{P}r\left(X > (1+\epsilon)\mu\right) + \mathbb{P}r\left(X < (1-\epsilon)\mu\right)
\end{aligned}$$

حال به محاسبه‌ی کران بالا برای $\mathbb{P}r\left(X > (1+\epsilon)\mu\right)$ می‌پردازیم. داریم:

$$\begin{aligned}
\mathbb{P}r\left(X > (1+\epsilon)\mu\right) &= \mathbb{P}r\left(e^{tX} > e^{t(1+\epsilon)\mu}\right) \\
&< \mathbb{E}[e^{tX}] \cdot e^{-t(1+\epsilon)\mu} && \text{(کران مارکوف)} \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \cdot e^{-t(1+\epsilon)\mu} && X = X_1 + \dots X_n \\
&= \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] \cdot e^{-t(1+\epsilon)\mu} \\
&= \prod_{i=1}^{n} (p_i \cdot e^t + (1-p_i)) \cdot e^{-t(1+\epsilon)\mu} \\
&= \prod_{i=1}^{n} (p_i(e^t - 1) + 1)) \cdot e^{-t(1+\epsilon)\mu} \\
&= e^{\sum_{i=1}^{n} p_i(e^t-1)} \cdot e^{-t(1+\epsilon)\mu} && 1 + x \leq e^x \\
&= e^{\mu(e^t - 1 - t - \epsilon t)} \\
&= \left(\frac{e^{\epsilon}}{(1+\epsilon)^{1+\epsilon}}\right)^{\mu} && \text{(قرار دهید } t = \ln(1+\epsilon)\text{)} \\
&= e^{\mu(\epsilon - (1+\epsilon)\ln(1+\epsilon))} \\
&\leq e^{\mu(\epsilon - (1+\epsilon)\frac{2\epsilon}{2+\epsilon})} && \left(\ln(1+\epsilon) \geq \frac{2\epsilon}{2+\epsilon}\right) \\
&= e^{-\frac{\epsilon^2\mu}{2+\epsilon}} \\
&\approx e^{-\Theta(\epsilon^2\mu)}
\end{aligned}$$

به‌طور مشابه و با قرار دادن $t = -\ln(1-\epsilon)$ خواهیم داشت:

$$\begin{aligned}
\mathbb{P}r\left(X < (1-\epsilon)\mu\right) &< \left(\frac{e^{-\epsilon}}{(1-\epsilon)^{1-\epsilon}}\right)^{\mu} \\
&\leq e^{-\frac{\epsilon^2\mu}{2}} \\
&\approx e^{-\Theta(\epsilon^2\mu)}
\end{aligned}$$

و در نتیجه با جمع دو احتمال خواهیم داشت:

$$\begin{aligned}
\mathbb{P}r\left(|X - \mathbb{E}X| > \epsilon\mu\right) &= \mathbb{P}r\left(X > (1+\epsilon)\mu\right) + \mathbb{P}r\left(X < (1-\epsilon)\mu\right) \\
&\leq e^{-\Theta(\epsilon^2\mu)} + e^{-\Theta(\epsilon^2\mu)} \\
&\approx 2 \cdot e^{-\Theta(\epsilon^2\mu)}
\end{aligned}$$

پس حکم برای حالت خاص متغیرهای تصادفی برنولی ثابت شد. $\square$

نکته. اگر توان دو برای ϵ در کران چرنوف نبود، کران بهتری را داشتیم چراکه:

$$\begin{aligned}
0 < \epsilon < 1 &\implies \epsilon^2 < \epsilon \\
&\implies \frac{\epsilon^2\mu}{3} < \frac{\epsilon\mu}{3} \\
&\implies -\frac{\epsilon\mu}{3} < -\frac{\epsilon^2\mu}{3} \\
&\implies e^{-\frac{\epsilon\mu}{3}} < e^{-\frac{\epsilon^2\mu}{3}} \\
&\implies 2e^{-\frac{\epsilon\mu}{3}} < 2e^{-\frac{\epsilon^2\mu}{3}}
\end{aligned}$$

پس کران‌بالای $2e^{-\frac{\epsilon\mu}{3}}$ کران‌بالای بهتری بود که متاسفانه ϵ^2 را در کران نامساوی چرنوف داریم.

نکته. از کران چرنوف نتیجه می‌شود که احتمال فاصله‌داشتن میانگین چند متغیر تصادفی از امیدریاضی میانگین آن‌ها با افزایش تعداد متغیرها کاهش می‌یابد. به طور دقیق‌تر اگر $X_1, \ldots, X_n$ متغیرهای تصادفی مستقلی باشند که $X_i \in [0, 1]$ ، $X = \sum_{i=0}^{n} X_i$، $Y = \frac{X}{n}$ و $\mu = \mathbb{E}X$ آنگاه خواهیم داشت:

$$\begin{aligned}
\mathbb{P}r(|Y - \mathbb{E}Y| > \epsilon\mathbb{E}Y) &= \mathbb{P}r(|\frac{X}{n} - \frac{\mathbb{E}X}{n} > \epsilon\frac{\mu}{n}) \\
&= \mathbb{P}r(|X - \mathbb{E}X > \epsilon\mu) \\
&\leq 2 \cdot e^{-\frac{\epsilon^2\mu}{3}} && \text{(کران چرنوف)} \\
&= 2 \cdot e^{-n\frac{\epsilon^2\mathbb{E}Y}{3}}
\end{aligned}$$

که با افزایش n و ثابت‌ماندن $\mathbb{E}Y$، $e^{-n\frac{\epsilon^2\mathbb{E}Y}{3}}$ کوچک می‌شود.

تا اینجا سه نامساوی احتمالاتی را دیدیم که کاربردهای فراوانی در ادامه خواهند داشت، برای اطلاعات بیشتر راجع به این کران‌ها می‌توانید از [?] استفاده کنید.

۴ شمارش تقریبی

حال مسئله‌ی شمارش تقریبی را که در [?] آمده‌است خواهیم دید و به بررسی راه حل‌هایی برای آن می‌پردازیم.

۱.۴ صورت مسئله

سؤال . الگوریتمی می‌خواهیم که دنباله‌ای از رویدادها را رصد کرده و در هر لحظه تعداد رویدادهایی که تا آن لحظه رخ داده‌است (و یا تخمینی از آن) را خروجی دهد. به طور دقیق‌تر، یک شمارنده، داده‌ساختاریست که تنها یک عدد n را نگهداری می‌کند و قادر به انجام ۳ عملیات زیر است:

- `init()`؛ مقدار عدد n را ۰ می‌گذارد.
- `update()`؛ عدد n را یک واحد افزایش می‌دهد، یعنی $n \leftarrow n + 1$.
- `query()`؛ عدد n و یا تخمینی از آن را خروجی می‌دهد.

در ابتدا عملیات `init` انجام می‌شود، یعنی $n = 0$ است. عملیات `update` نشان‌دهنده‌ی رخ‌دادن یک رویداد جدید است و عملیات `query` تعداد رویدادها رخ‌داده تا آن لحظه را خروجی می‌دهد.

الگوریتم بدیهی نیاز به $O(\log n)$ بیت حافظه دارد؛ در ادامه ثابت می‌کنیم که حل مسئله به صورت قطعی با $o(\log n)$ بیت غیر ممکن است. پس هدف ما طراحی عملیات `query()` به صورتی است که عدد $\tilde{n}$ که تخمینی از n است را با حافظه‌ی کمتر خروجی دهد.

۲.۴ حداقل میزان حافظه برای خروجی قطعی

لم زیر ثابت می‌کند در حالتی که تابع `query()` مقدار دقیق تعداد اجرای `update()` را خروجی دهد، به لاآقل $\lceil \lg n \rceil$ بیت حافظه نیاز داریم که n تعداد دفعات اجرای `update()` است.

لم ۱۰. اگر شمارنده‌ی A دارای k بیت از حافظه باشد، در خروجی عملیات `query()` در حداقل یکی از $2^k + 1$ بار اجرای `query()` دقیقاً پس‌از اجراشدن عملیات `update()` خطا دارد.

اثبات. با توجه به اینکه حافظه‌ی ماشین k بیت دارد، پس تعداد حالت‌های متفاوت در حافظه‌ی ماشین 2^k تاست، پس از $2^k + 1$ بار اجرای تابع `update()` حداقل دو بار حالت ماشین یکسان می‌شود، درنتیجه خروجی عملیات `query()` نیز در این دو حالت یکسان خواهد شد، ولی n در این دو اجرا متفاوت است، در نتیجه خروجی حداقل یکی از این دو عملیات `query()` نادرست خواهد بود. $\square$

۳.۴ در ادامه

تا اینجا دیدیم که برای حالت قطعی، کران پایین $\Omega(\lg n)$ را خواهیم داشت، اما می‌توان با استفاده از تقریب، با کمک حافظه‌ی کمتری - مثلاً $\lg \lg n$- شمارنده ساخت. در جلسه‌ی بعدی به بررسی بیشتر مسئله‌ی شمارش تقریبی و روش‌های ساخت آن با حافظه‌ی کمتر خواهیم پرداخت.