جمع متغیرهای تصادفی مستقل

امید اعتصامی
پژوهشگاه دانشهای بنیادی، پژوهشکده ریاضیات

۱ انگیزه

در بسیاری از مسائل ما علاقه‌مند به متغیر تصادفی هستیم که خود جمع چند متغیر تصادفی دیگر است. به عنوان مثال در گردش تصادفی [۱]، ما در هر مرحله مستقل از مراحل قبلی موقعیت خود را تغییر می‌دهیم و موقعیت ما در هر لحظه جمع چند متغیر تصادفی مستقل است.

خیلی اوقات می‌توان برای چنین متغیر تصادفی نشان داد که با احتمال خیلی کمی خیلی بزرگ یا خیلی کوچک است.

۲ قضیه‌ی حد مرکزی

قضیه ۱ فرض کنید X_1، X_2، ... دنباله‌ای از متغیرهای تصادفی مستقل با توزیع یکسان، دارای متوسط متناهی μ و واریانس متناهی σ^2 باشند. آنگاه توزیع

$$\frac{(X_1 + \ldots + X_n) - n\mu}{\sigma\sqrt{n}}$$

به توزیع متغیر نرمال $N(0,1)$ دارای متوسط ۰ و واریانس ۱ میل می‌کند.

به یاد آورید که متغیر نرمال $N(0,1)$ دارای تابع چگالی احتمال

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

می‌باشد.

۳ کومولان

ما برای قضیه‌ی حد مرکزی اثبات کامل ارائه نمی‌کنیم، اما سعی می‌کنیم از طریق کومولان‌ها [۲] توضیح دهیم که چرا می‌تواند درست باشد.

تابع مولد کومولان متغیر تصادفی X برابر با

$$K(\theta) = \ln(\mathbb{E}e^{\theta X})$$

تعریف می‌شود. اگر بسط تیلور $K(\theta)$ را حول $\theta = 0$ بنویسیم، یعنی

$$K(\theta) = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!}\theta^n,$$

آنگاه به κ_n کومولان nام می‌گویند. به عبارت دیگر κ_n مشتق nام تابع مولد کومولان در نقطه‌ی $\theta = 0$ است.

می‌توان دید که κ_1 همان امید ریاضی X و κ_2 همان واریانس X می‌باشد.

[۱] random walk
[۲] cumulants

۴ خواص کومولان

تابع مولد متغیر تصادفی نرمال $N(0,1)$ برابر است با

$$\begin{aligned} K(\theta) &= \ln(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} e^{\theta x} dx) \\ &= \ln(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2} e^{\theta^2/2} dx) \\ &= \ln(e^{\theta^2/2}) = \theta^2/2. \end{aligned}$$

بنابراین برای $N(0,1)$ تمام کومولان‌ها صفر هستند بجز کومولان دوم که برابر یک است.

از طرف دیگر اگر X و Y متغیرهای تصادفی مستقل باشند، کومولان nام $X+Y$ برابر مجموع کومولان nام X و Y می‌باشد:

$$\ln(\mathbb{E}e^{\theta(X+Y)}) = \ln((\mathbb{E}e^{\theta X})(\mathbb{E}e^{\theta Y})) = \ln(\mathbb{E}e^{\theta X}) + \ln(\mathbb{E}e^{\theta Y}).$$

و اگر متغیر تصادفی X را ضرب در ثابت c کنیم، آنگاه کومولان nام آن ضرب در c^n می‌شود.

و همچنین اگر متغیر تصادفی X را با عدد c جمع کنیم، کومولان nام به ازای $n \geq 2$ تغییر نمی‌کند. (به همین خاطر گاهی به کومولان‌ها semi-invariants هم می‌گویند.)

بنابراین در قضیه‌ی حد مرکزی، اگر کومولان kام X_iها برابر κ_k باشد، آنگاه کومولان اول

$$X = \frac{(X_1 + \ldots + X_n) - n\mu}{\sigma\sqrt{n}}$$

برابر صفر، و کومولان kام X به ازای $k \geq 2$ برابر $\frac{n\kappa_k}{\sigma^k n^{k/2}}$ خواهد بود. یعنی به ازای nهای بزرگ کومولان دوم به یک و کومولان‌های بالاتر به صفر میل می‌کنند. پس نشان دادیم که کومولان‌های X به کومولان‌های $N(0,1)$ میل می‌کنند.

۵ نابرابری چرنوف برای متغیرهای همگن

صورتهای مختلفی برای نابرابری چرنوف [۳] وجود دارد. ما بعضی از آنها را در این جا می‌آوریم.

قضیه ۲ فرض کنید $X = X_1 + \ldots + X_n$ که در آن متغیرهای تصادفی X_1، …، X_n مستقل و دودویی هستند و $\Pr(X_i = 1) = p$. قرار دهید $q = 1 - p$. آنگاه به ازای $\epsilon \geq 0$ نابرابری زیر برقرار است:

$$\Pr(X \geq n(p+\epsilon)) \leq e^{-D(p+\epsilon||p)n}$$

که در آن

$$D(p+\epsilon||p) = (p+\epsilon)\ln\frac{p+\epsilon}{p} + (q-\epsilon)\ln\frac{q-\epsilon}{q}.$$

همچنین نتایج زیر از نابرابری بالا بدست می‌آیند:

$$\Pr(X \geq n(p+\epsilon)) \leq e^{-\epsilon^2 n/(2q)}, \quad \Pr(X \geq n(p+\epsilon)) \leq e^{-2\epsilon^2 n}$$

و اگر $p \geq 1/2$، آنگاه

$$\Pr(X \geq n(p+\epsilon)) \leq e^{-\epsilon^2 n/(2pq)}.$$

اثبات: دو روش برای اثبات این نابرابری وجود دارد. روش اول بر پایه‌ی نابرابری مارکوف

$$\Pr(X \geq r) \leq \mathbb{E}e^{tX}/e^{tr}.$$

[۳] Chernoff

و بهینه سازی پارامتر t می‌باشد. اما روش دوم بر پایه‌ی این مشاهده است: X' را مشابه X برابر $X_1' + \ldots + X_n'$ تعریف کنید با این تفاوت که $\Pr(X_i' = 1) = p + \epsilon$. آنگاه به ازای $k \geq n(p+\epsilon)$ داریم

$$\Pr(X' = k)/\Pr(X = k) \geq (\frac{p+\epsilon}{p})^{n(p+\epsilon)}(\frac{q-\epsilon}{q})^{n(q-\epsilon)}.$$

چون $\Pr(X' \geq n(p+\epsilon)) \leq 1$ نابرابری بدست می‌آید. □

از نابرابری‌های بالا می‌توان برای $\Pr(X \leq n(p-\epsilon))$ هم کران بالا بدست آورد. مثلا

$$\Pr(X \leq n(p-\epsilon)) = \Pr(n - X \leq n(q+\epsilon)) \leq e^{-\epsilon^2 n/(2p)}.$$

همچنین توجه کنید که ضریب $-D(p+\epsilon||p)$ در توان بهینه است چون احتمال این که $X = \lceil n(p+\epsilon) \rceil$ (طبق تخمین استرلینگ برای فاکتوریل) حداقل برابر

$$2^{-D(p+\epsilon||p)n+o(n)}$$

می‌باشد.

۶ کم کردن احتمال خطای الگوریتم‌ها

فرض کنید مساله‌ای وجود دارد که جواب آن یا صفر یا یک است. فرض کنید الگوریتمی داشته باشیم که جواب مساله را با احتمال ۲/۳ درست حدس بزند و این الگوریتم را n بار تکرار کنیم.

اگر X_i متغیر تصادفی نشانگر باشد که برابر ۱ باشد اگر الگوریتم جواب غلط بدهد و در غیر اینصورت $X_i = 0$، آنگاه $\Pr(X_i = 1) = 1/3$. حال طبق نابرابری چرنوف

$$\Pr(X_1 + \ldots + X_n \geq n/2) \leq e^{-2\cdot(1/6)^2 n} = e^{-n/18}.$$

پس اگر ما اکثریت جواب الگوریتم را در n بار تکرار الگوریتم به عنوان جواب مساله بدهیم، احتمال خطای ما حداکثر برابر است با $e^{-n/18}$، یعنی به صورت نمایی کم شده است.

۷ احتمال خطای نمونه‌برداری

فرض کنید که به شما یک فرمول DNF ϕ داده شده است. یکی از مسائل اساسی در نظریه‌ی محاسبات پیدا کردن احتمال صدق‌پذیری[۴] فرمول ϕ وقتی است که متغیرهای دودویی ورودی فرمول ϕ هر کدام تصادفی یکنواخت مستقل باشند. برای محاسبه‌ی این احتمال می‌توان از روش نمونه‌برداری استفاده کرد: یک ورودی کاملا تصادفی به فرمول ϕ داد و دید که آیا ϕ صدق پیدا کرد یا نه. آنگاه این کار را n بار انجام داد و کسری از n بار که ϕ در آن صدق پیدا کرده بود را به عنوان تخمین احتمال صدق‌پذیری ارائه کنیم. می‌توان با نابرابری چرنوف نشان داد که اگر

$$n = \Theta(\epsilon^{-2}\log(1/\delta))$$

آنگاه خطای تخمین ما به احتمال $1-\delta$ حداکثر ϵ است.

۸ نابرابری چرنوف به صورت ضربی

قضیه ۳ فرض کنید $X = X_1 + \ldots + X_n$ که در آن متغیرهای تصادفی X_1، ...، X_n مستقل و دودویی هستند و $\Pr(X_i = 1) = p_i$. تعریف کنید $\mu = \mathbb{E}X$. آنگاه

$$\Pr(X \geq r\mu) \leq (r^{-r}e^{r-1})^{\mu} \quad if\, r \geq 1$$

و

$$\Pr(X \leq r\mu) \leq (r^{-r}e^{r-1})^{\mu} \quad if\, 0 < r \leq 1.$$

[۴] satisfying probability

فرم ساده‌تر نابرابریهای بالا به ازای $0 \leq \delta \leq 1$ عبارتند از

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\delta^2\mu/3}, \quad \Pr(X \geq (1-\delta)\mu) \leq e^{-\delta^2\mu/2}.$$

۹ میانه متغیر دوجمله‌ای

قضیه‌ی زیر نشان می‌دهد که در متغیرهای دوجمله‌ای (مجموع متغیرهای مستقل دودویی با توزیع یکسان)، امید ریاضی متغیر تصادفی با مقدار میانه‌ی متغیر تصادفی فرقی ندارد.

قضیه ۴ فرض کنید $X = X_1 + \ldots + X_n$ که در آن متغیرهای تصادفی X_1، …، X_n مستقل و دودویی هستند و $\Pr(X_i = 1) = m/n$ که در آن m عدد صحیحی است و $0 \leq m \leq n$. آنگاه m میانه X است، یعنی

$$\Pr(X < m) < 1/2, \quad \Pr(X > m) < 1/2.$$

۱۰ نابرابری هفدینگ

هفدینگ نشان داد که اگر X مجموع تعدادی متغیر دودویی مستقل باشد، احتمال دور بودن از مقدار میانگین موقعی بیشتر است که تمام متغیرها توزیع یکسان داشته باشند.

قضیه ۵ فرض کنید $X = X_1 + \ldots + X_n$ و $X' = X'_1 + \ldots + X'_n$ که در آن متغیرهای تصادفی X_1، …، X_n و X'_1، …، X'_n مستقل و دودویی هستند و $\Pr(X'_i) = \mathbb{E}X/n$. آنگاه

$$\Pr(X' \leq m) \leq \Pr(X \leq m) \textit{ if } m \leq \mathbb{E}X - 1, \quad \Pr(X' \geq m) \leq \Pr(X \geq m) \textit{ if } m \geq \mathbb{E}X + 1.$$

۱۱ جفت کردن دو متغیر تصادفی

قضیه ۶ فرض کنید $X = X_1 + \ldots + X_n$ که در آن متغیرهای تصادفی X_1، …، X_n مستقل و دودویی هستند و $\Pr(X_i) = p$. آنگاه به ازای m صحیح بین ۰ و n، داریم

$$\Pr(X \leq m) = (n-m)\binom{n}{m}\int_{x=p}^{1} x^m(1-x)^{n-m-1}dx.$$

اثبات: برای اثبات این قضیه از روش جالب جفت کردن [۵] استفاده می‌کنیم. فرض کنید که n عدد تصادفی مستقل یکنواخت بین ۰ و ۱ انتخاب می‌کنیم و آنها را Y_1، …، Y_n می‌نامیم. حال طبق نماد ایورسون تعریف کنید

$$X_i^{(x)} = [Y_i \leq x].$$

به وضوح می‌توان به جای X_i از $X_i^{(p)}$ استفاده کرد که همان توزیع را دارد. از طرفی

$$\Pr(\sum_i X_i^{(x)} \leq m) - \Pr(\sum_i X_i^{(x+dx)} \leq m)$$

در حد وقتی $dx \to 0$ برابر احتمال این است که m تا از Y_1، …، Y_n در بازه‌ی $[0, x]$ باشند و تعدادی از آنها در بازه‌ی $(x, x+dx]$ باشند. $\binom{n}{m}$ حالت برای انتخاب متغیرهای داخل بازه‌ی $[0, x]$ وجود دارد و چون dx بی‌نهایت کوچک است می‌توان فرض کرد که دقیقا یکی از $n-m$ متغیر دیگر در بازه‌ی $(x, x+dx]$ وجود دارد. پس

$$\Pr(\sum_i X_i^{(x)} \leq m) - \Pr(\sum_i X_i^{(x+dx)} \leq m) = (n-m)\binom{n}{m}x^m(1-x)^{n-m-1}dx.$$

□

[۵] coupling

۱۲ یک نابرابری دور از انتظار

قضیه ۷ فرض کنید X_1، X_2، ...، متغیرهای تصادفی مستقل با توزیع یکسان گسسته باشند. آنگاه

$$\Pr(|\frac{X_1+\ldots+X_{2n}}{2n}-\alpha| \leq |\frac{X_1+\ldots+X_n}{n}-\alpha|) > 1/2.$$

اثبات آن را به عنوان تمرین برای خواننده می‌گذاریم. آن چه در باره‌ی این نابرابری دور از انتظار است این است که به ازای هر مقدار α و نه فقط $\alpha = \mathbb{E}X_1$ درست است.

۱۳ یک مساله‌ی باز

با وجود این که خیلی چیزها درباره‌ی جمع متغیرهای مستقل می‌دانیم، بعضی سوالات ساده در مورد آنها هم چنان باز هستند.

حدس ۸ اگر X_1، ...، X_n متغیرهای تصادفی مستقل نامنفی هر یک دارای امید ریاضی برابر ۱ باشند، آنگاه

$$\Pr(X_1+\ldots+X_n < n+1/(e-1)) \geq 1/e.$$

این حدس را فایگه [۶] در رابطه با تخمین میانگین درجه‌ی گرافها زده است [۱].

مراجع

[1] Uriel Feige. (2004) On sums of independent random variables with unbounded variance, and estimating the average degree in a graph. Proceedings of the 36th Symposium on Theory of Computing, pages 594-603.

[۶] Feige