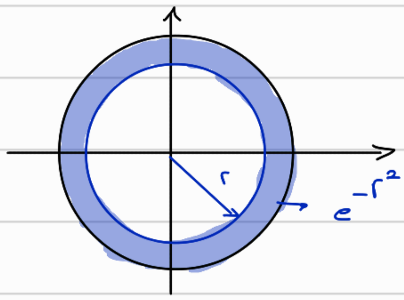


$$\int e^{-\|x\|_2^2}\, dx_1 \dots dx_d = \int_0^\infty \left(\text{در ناحیه‌های مختلف در فاصله‌ی } r\right) dr$$

$e^{-r^2} . dr . A_d(r)$

$$= A_d \int_0^\infty e^{-r^2} r^{d-1}\, dr$$

$x = r^2$

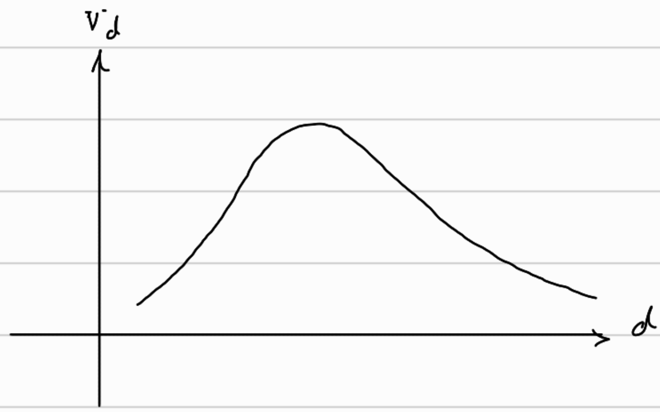
$$= A_d \int_0^\infty e^{-x} x^{\frac{d-1}{2}} \frac{dx}{2\sqrt{x}}$$

$$= \frac{A_d}{2} \int_0^\infty e^{-x} x^{\frac{d}{2}-1}\, dx$$

$$= \frac{\Gamma(\frac{d}{2})}{2} A_d$$

یادآوری: $\Gamma(t) = \int_0^\infty e^{-x} x^{t-1}\, dx$

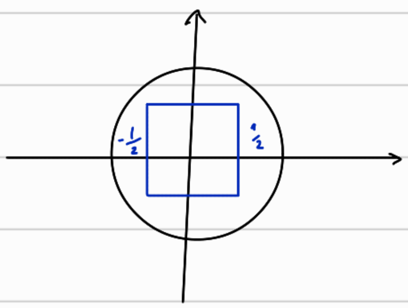
$$\Rightarrow \boxed{A_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}} \Rightarrow \boxed{V_d = \frac{2}{d} \frac{\pi^{d/2}}{\Gamma(d/2)}}$$



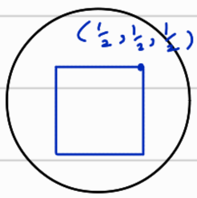
$$\lim_{d \to \infty} V_d = 0$$

مقایسه با مکعب با ضلع واحد:

$d=2$



$d=2$



همچنان کل مکعب داخل کره است

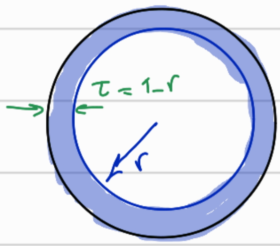
$d=4$

در بعد بالا نقاط گوشه مکعب بیرون از کره می‌افتند. امّا دقت کنید کره همیشه شامل نقطه‌ی $(0, \dots, 0, 0, 1)$ می‌شود که خارج مکعب است پس کره هیچگاه زیر مجموعه مکعب نمی‌شود.

$V = \left(\frac{2}{\sqrt{d}}\right)^d$

طول ضلع مکعب محاطی: $\frac{2}{\sqrt{d}}$ ، طول ضلع مکعب محیطی: 2 ← کره شبیه‌تر به مکعب محاطی است

* آیا حجم کره در نزدیک پوسته بیشتر است یا مرکز؟

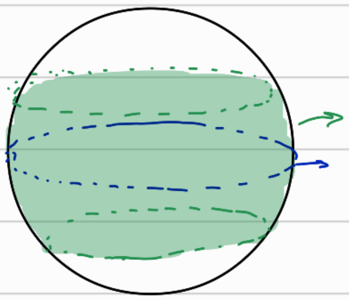


$$\text{حجم هاشور خورده} = V_d(1) - V_d(r)$$

$$\frac{\text{حجم هاشور خورده}}{\text{کل حجم}} = 1 - \frac{V_d(r)}{V_d(1)} = 1 - r^d = 1-(1-\tau)^d \overset{\tau \ll 1}{\approx} 1 - e^{-\tau d}$$

* به ازای چه τ : نسبت $= \frac{2}{3}$ ؟ $\Rightarrow 1 - e^{-\tau d} = \frac{2}{3} \Rightarrow e^{-\tau d} = \frac{1}{3} \Rightarrow \tau = \frac{1}{d}\log 3$

نکته :

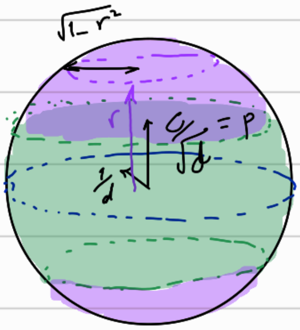


در بعد بالا کل حجم کره اطراف دایره عظیمه جمع می شود

نکته بالا در جلسه آینده ثابت خواهد شد.

جلسه سوم 1403, 7, 8

* ادامه مبحث تمرکز اندازه برای کره



$$\frac{V(green)}{V_d} \geq 1 - \epsilon$$

$$V(purple) = \int_p^1 \overbrace{V_{d-1}\cdot(1-r^2)^{\frac{d-1}{2}}}^{V_{d-1}(\sqrt{1-r^2})} dr$$

$$\Rightarrow \frac{V(purple)}{V_d} = \frac{V_{d-1}}{V_d}\int_p^1 (1-r^2)^{\frac{d-1}{2}} dr$$

یادآوری : $1 - x \leq e^{-x}$, $1 - nx \leq (1-x)^n \leq e^{-nx}$: اگر $0 \leq x \leq 1$

$$\Rightarrow \frac{V(purple)}{V_d} \leq \frac{V_{d-1}}{V_d}\int_p^1 e^{-\frac{(d-1)}{2}r^2} dr \leq \frac{V_{d-1}}{V_d}\int_p^\infty e^{-(d-1)\frac{r^2}{2}} dr$$

یادآوری : $\int_p^\infty e^{-\frac{(d-1)}{2}r^2} dr < C \cdot e^{-\frac{d-1}{2}p^2}$

یادآوری : $\int_p^\infty e^{-r^2\beta} dr < \int_p^\infty \frac{r}{p} e^{-r^2\beta} dr \overset{u=r^2\beta}{=} \int_{p^2\beta}^\infty e^{-u} du \, \frac{1}{2p\beta} = e^{-p^2\beta}\frac{1}{2p\beta}$

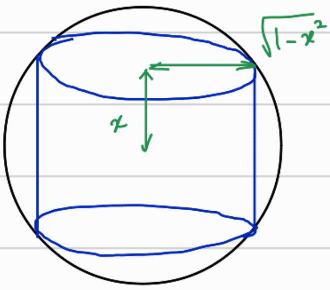
$$\Longrightarrow \frac{V(purple)}{V_d} \le \frac{V_{d-1}}{V_d} \, \frac{e^{-p^2\frac{d-1}{2}}}{p(d-1)}$$

حال ثابت می‌کنیم $\frac{V_{d-1}}{V_d} < \sqrt{d-1}$. درنتیجه با این فرض داریم

$$\frac{V_{d-1}}{V_d} < \sqrt{d-1} \; ; \; \frac{V(purple)}{V_d} \le \frac{1}{p\sqrt{d-1}} e^{-\frac{p^2}{2}(d-1)} \overset{\text{اگر } p=\frac{c}{\sqrt{d-1}}}{=} \frac{1}{c} e^{-\frac{c^2}{2}} \overset{\text{اگر } c>1}{<} e^{-\frac{c^2}{2}}$$

$$\Longrightarrow \frac{V(green)}{V_d} \ge 1-2e^{-\frac{c^2}{2}}$$

اثبات رابطه‌ی $\frac{V_{d-1}}{V_d}$:



$$V(\text{استوانه}) < V_d$$

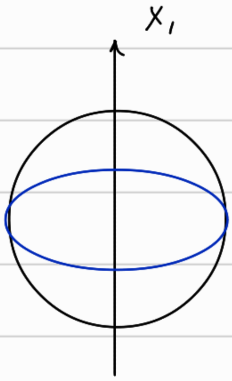
$$2x \, V_{d-1} (1-x^2)^{\frac{d-1}{2}} < V_d$$

$$\frac{V_{d-1}}{V_d} < \frac{1}{2x(1-x^2)^{\frac{d-1}{2}}} \overset{x=\frac{1}{\sqrt{d-1}}}{\Longrightarrow} \frac{V_{d-1}}{V_d} < \frac{1}{2x\left(1-\frac{d-1}{2}x^2\right)} = \frac{1}{2\frac{1}{\sqrt{d-1}}\frac{1}{2}} = \sqrt{d-1}$$

قضیه ۱: یک نقطه تصادفی و یکنواخت از درون کره واحد انتخاب می‌کنیم.

$$\underbrace{Pr(R<1-\epsilon)}_{\frac{V_d(\text{کره به شعاع } 1-\epsilon)}{V_d(1)}} < e^{-d\epsilon}$$

قضیه 2 :



$$X=(X_1,\dots,X_d)$$

$$Pr\left(|X_1| \ge \frac{c}{\sqrt{d-1}}\right) \le 2e^{-\frac{c^2}{2}} \; ; \; c>1$$

یا $Pr(\underbrace{\sqrt{d-1}\,|X_1|}_{Y_1} \ge c) \le 2e^{-\frac{c^2}{2}}$

$$\Pr\left(\forall i : |X_i| < \frac{c}{\sqrt{d-1}}\right) = 1 - \Pr\Big(\underbrace{\exists i : |X_i| > \frac{c}{\sqrt{d-1}}}_{\bigcup \left\{ |X_i| > \frac{c}{\sqrt{d-1}} \right\}}\Big)$$

$$\geq 1 - \sum_i \Pr\left(|X_i| > \frac{c}{\sqrt{d-1}}\right) \geq 1 - 2d\, e^{-\frac{c^2}{2}}$$

$$\xrightarrow[\text{آمدن جمله دوم}]{\text{برای ثابت}} \quad c = O\left(\sqrt{\ln d}\right) \Longrightarrow \|X\|_\infty = \Omega\left(\sqrt{\frac{\ln d}{d}}\right) = \tilde{\Omega}\left(\frac{1}{\sqrt{d}}\right)$$

* در بعد بالا (d) می‌توان 2^{cd} بردار متعامد تقریبی داشت.

اثبات (روش احتمالاتی): فرض کنید $v_1, \dots, v_M$ ، M بردار تصادفی یکنواخت و مستقل از روی کره باشند. نشان می‌دهیم

$$\Pr\left[\exists i, j : |\langle v_i, v_j \rangle| > \epsilon\right] \leq 0.99$$

$$E\langle v_i, v_j\rangle = E\, v_i^T v_j \overset{\text{استقلال}}{=} E\, v_i^T . E\, v_j = 0$$

روش اثبات:
$$\Pr\left[\exists (i,j) : \dots\right] \leq \sum_{(i,j)} \Pr\left[|\langle v_i, v_j\rangle| > \overset{c/\sqrt{d-1}}{\epsilon}\right] < 2 \sum_{(i,j)} e^{-c^2/2} = M^2 e^{-c^2/2} < 0.99$$

جلسه چهارم 9، 7، 1403

ادامه اثبات جلسه قبل:

$$M^2 e^{-c^2/2} \leq 0.99 \Longrightarrow M \leq e^{c^2/4}\sqrt{0.99} \xRightarrow{c \approx \sqrt{d}\,\epsilon} M \leq \exp(K . d . \epsilon^2) = 2^{O(d)}$$

* فصل 2 ورشنین
* نامساوی‌های اولیه
* قوانین حدی
* نامساوی چرنوف برای متغیرهای برنولی و ارتباط با پواسون
* نامساوی‌های اولیه

(ینسن) فرض کنید $\varphi : \mathbb{R}^d \to \mathbb{R}$ یک تابع محدب

$$E(\varphi(X)) \geq \varphi(EX) \xrightarrow{\text{مثال}} \begin{cases} E(X^2) \geq E^2(X) \\ E\left(\frac{1}{X}\right) \geq \frac{1}{EX} \end{cases}$$

نرم متغیر تصادفی و ضرب داخلی :

فرض کنید X و Y دو متغیر تصادفی وابسته باشند.

$* \; \langle X, Y \rangle = E[XY]$

$* \; \|X\|_{L^2} = \sqrt{E(X^2)} \quad , \quad \|X\|_{L^p} = \left(E|X|^p\right)^{1/p}$

$* \; \|u\|_2 = \max_{\|v\|_2 = 1} \langle u, v \rangle \Rightarrow \|X\|_{L^2} = \max_{Y: EY^2 = 1} E(XY)$

$* \; \|X\|_{L^p} = \max_{Y: \|Y\|_{L^q} = 1} E[XY] \quad , \quad \frac{1}{p} + \frac{1}{q} = 1$

مزدوج هولدر

نامساوی هولدر : $E(XY) \leq \|X\|_{L^p} \|Y\|_{L^q} \quad ; \quad p, q > 0$

اثبات : $E\,XY \leq (\underbrace{E|X|^p}_{v})^{1/p} \cdot (\underbrace{E|Y|^q}_{v'})^{1/q} \Rightarrow E\, v^{1/p} v'^{1/q} \leq (E v)^{1/p} (E v')^{1/q}$

می‌توان نشان داد $f(u, v) = u^{1/p} v^{1/q}$ مقعر است پس نامساوی بالا درست است.

$* \; \|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$

$$\sup_{Z: \|Z\|_{L^q} = 1} E\left(Z(X+Y)\right) = \sup \left(E(ZX) + E(ZY)\right) \leq \sup_{Z} E(ZX) + \sup_{\tilde{Z}} E(ZY) = \|X\|_{L^p} + \|Y\|_{L^p}$$

* اگر $1 \leq p < q$ باشد $\|X\|_{L^p} \leq \|X\|_{L^q}$

فضاهای حدی :

* قانون اعداد بزرگ (LLN) : فرض کنید $X_1, \ldots, X_n$ متغیرهای تصادفی iid با میانگین μ باشند.

$$\frac{\sum X_i}{n} \xrightarrow[n \to \infty]{} \mu \quad \text{یا} \quad \Pr\left[\frac{\sum X_i}{n} \notin (\mu \pm \delta)\right] \xrightarrow[n \to \infty]{} 0$$

در احتمال

* قانون حد مرکزی (CLT) : اگر $X_1, \ldots, X_n$ متغیرهای تصادفی iid باشند با میانگین صفر و $\mathrm{Var}(X_i) < \infty$ آنگاه

$$\forall (a,b):\quad \underbrace{Pr\left[\frac{\sum X_i}{\sqrt{n}} \in (a,b)\right]}_{A} \xrightarrow[n\to\infty]{} \underbrace{\int_a^b \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma^2}}\, dx}_{B}$$

به عبارت دقیق‌تر: $\sup_{(a,b)} |A-B| \xrightarrow[n\to\infty]{} 0$

* نامساوی چرنوف:

مارکوف: $Pr[X>t] \le \frac{E(X)}{t}$; $X \geq 0$

چبیشف: $Pr(|X-\mu|>t) \le \frac{Var}{t^2}$

چرنوف: اگر $s>0$: $P(X>t) \le e^{-st} E(e^{sX})$

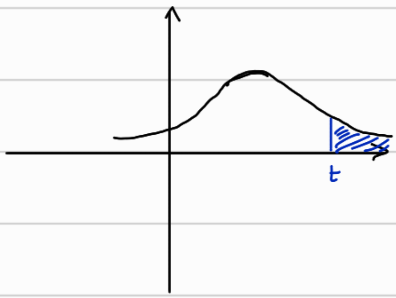
$e^{sX} \geq e^{st}$

جلسه پنجم 15, 7, 1403

* نامساوی چرنوف برای متغیرهای باینری
* مثال گراف های تصادفی
* متغیرهای زیرگوسی

* نامساوی چرنوف:

$$P(X\geq t) \le e^{-st}\, \overbrace{E[e^{sX}]}^{\varphi_X(s)} \quad ; \forall s>0$$



$$P(|X|\geq t) \le \frac{E[|X|^q]}{t^q} \quad ; \forall q \to$$ اثبات مانند چرنوف

اگر $X_1, \dots, X_n$ متغیرهای تصادفی مستقل باشند

$$\varphi_{X_1+\dots+X_n}(s) = E[e^{s\sum X_i}] = E\left[\prod e^{sX_i}\right] = \prod \varphi_{X_i}(s) \to \text{Tensorization}$$

نکته (تمرین): $\inf_q \frac{E[|X|^q]}{t^q} \le \inf_s e^{-st} E[e^{sX}]$

فرض کنید $X_i \sim Ber(p_i)$ باشند و $S = \sum_{i=1}^n X_i$ و $\sum_{i=1}^n p_i = \mu$. هدف باند زدن روی $Pr[S > \overbrace{(1+\delta)\mu}^{t}]$ است.

$$Pr[S>t] \leq e^{-\lambda t} E[e^{\lambda S}] = e^{-\lambda t}\prod_{i=1}^{n} E[e^{\lambda X_i}]$$

$$= e^{-\lambda t}\prod_{i=1}^{n}\left(\overbrace{p_i e^{\lambda} + (1-p_i)}^{1+p_i(e^{\lambda}-1)\ \leq\ e^{p_i(e^{\lambda}-1)}}\right)$$

$$\leq e^{-\lambda t}\prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)} = e^{-\lambda t}\, e^{\mu(e^{\lambda}-1)}$$

یادآوری: $Y\sim Poi(\mu) \Rightarrow \varphi_x(\lambda) = e^{\mu(e^{\lambda}-1)}$ → یعنی MGF برای S از یک متغیر تصادفی پواسون کمتر است

با بهینه‌سازی روی λ به این نتیجه می‌رسیم که $\lambda = \ln\frac{t}{\mu}$ و درنتیجه:

$$P(S\geq t) \leq e^{-\mu}\cdot\left(\frac{e\mu}{t}\right)^{t} \Rightarrow P(S\geq(1+\delta)\mu) \leq e^{\delta\mu}\left(\frac{1}{1+\delta}\right)^{\mu(1+\delta)} \overset{\text{تمرین}}{\leq} \exp(-c_R\cdot\mu\cdot\delta^2)$$

به شکل مشابه می‌توان نوشت

$$P(S\leq(1-\delta)\mu) \leq e^{\delta\mu}\left(\frac{1}{1-\delta}\right)^{\mu(1-\delta)} \leq \exp(-c_L\cdot\mu\cdot\delta^2) \ ; \ 0<\delta<1$$

گراف‌های تصادفی اردوش-رنیی: $G(n,p)$

وجود یال با احتمال p
و مستقل از بقیه یال‌ها

این گراف برای p های به اندازه کافی بزرگ ولی نه چندان بزرگ، گراف تقریباً منتظم (δ-منتظم) است.

درجه‌ی رأس i ام

$$\forall i,j: \ 1-\delta \leq \frac{d_i}{d_j} \leq 1+\delta$$

فرض کنید درجه‌ی رأس‌های گراف $(d_1,\dots,d_n)$ باشند.

$$d_i \sim Bin(n-1,p)$$

$$d_i = \sum 1\left\{\begin{matrix}\text{یال بین } j,i\\ \text{وجود دارد}\end{matrix}\right\} \Rightarrow \bar{d} = E[d_i] = (n-1)p$$

خاصیت δ-منتظم: $d_i \in (1\pm\delta)\bar{d} \ : \forall i$

هدف: هر گراف $G(n,p)$ برای p های خوب تقریباً منتظم است.

$$Pr[\exists i: d_i\notin(1\pm\delta)\bar{d}] \leq \sum_i Pr[d_i\notin(1\pm\delta)\bar{d}] = n\cdot Pr[d_i\notin\dots]$$

$$\leq n \cdot 2\exp(-c \cdot \bar{d} \cdot \delta^2)$$

$$= \underbrace{2n\exp(-c \cdot (n-1)p \cdot \delta^2)}_{o(\frac{1}{n^2})} = o(\tfrac{1}{n})$$

$$\Rightarrow \log \frac{1}{n^2} = -c \cdot \delta^2 \cdot (n-1)p \Rightarrow p \gtrsim \frac{\log n}{n \cdot \delta^2}$$

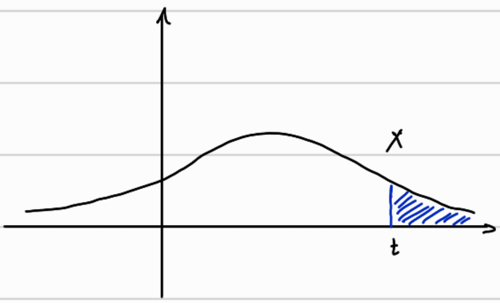
* تمرین های کتاب (Vershynin)

- $p = O(\frac{\log n}{n})$ ← همه‌ی درجات از مرتبه $O(\log n)$ هستند.
- $p = O(\frac{1}{n})$ ← $\max d_i = O\left(\frac{\log n}{\log \log n}\right)$
- $p = o(\frac{\log n}{n})$ ← حداقل یک رأس از درجه‌ی $10\bar{d}$ دارد.

جلسه ششم $1403, 7, 17$

* متغیرهای زیرگوسی
* تعریف
* خواص
* قضیه هوفدینگ
* لم هوفدینگ متغیرکران‌دار

* جلسه قبل: $S = \sum X_i, \; X_i \sim Ber(p_i) \Rightarrow \varphi_S(\lambda) \leq \varphi_{Poi(\mu)}(\lambda), \quad \mu = \sum p_i$



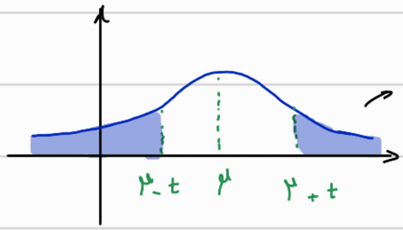
* متغیر تصادفی X، زیر (Z) است اگر $\varphi_X(\lambda) \leq \varphi_Z(\lambda) \; ; \; \forall \lambda$.

کاربرد زیر (Z) بودن: $\Pr(X > t) \leq e^{-\lambda t}\varphi_X(\lambda) \leq e^{-\lambda t}\varphi_Z(\lambda)$

تعریف زیرگوس: X را σ^2-زیرگوسی گوییم اگر $\varphi_X(s) \leq \underbrace{e^{\frac{s^2\sigma^2}{2}}}_{\text{مولد گشتاور } \mathcal{N}(0,\sigma^2)}$ (Vershynin)

حالت کلی‌تر: $\mathbb{E}\left[e^{s(X-\mu)}\right] \leq e^{\frac{s^2\sigma^2}{2}}$ (هم مراجع دیگر) $(\forall s)$

* نتیجه چرنوف و تعریف: $P(X > t) \leq e^{-st}\varphi_X(s) \leq e^{-st + \frac{s^2\sigma^2}{2}} \xRightarrow{\inf_s} P(X > t) \leq e^{-\frac{t^2}{2\sigma^2}}$



$e^{-\frac{t^2}{2\sigma^2}}$

بانو چبیشف

$$P(|X-\mu|>t) \le 2e^{-\frac{t^2}{2\sigma^2}} \leq \frac{\sigma^2}{t^2}$$

* یک متغیر σ^2 زیرگوسی دارای خاصیت زیر است: $Var(X) \le \sigma^2$ (تمرین)

* خواص معادل برای بررسی زیرگوسی بودن: (فرض کنیم میانگین X صفر است)

① $E[e^{sX}] \le e^{\frac{s^2 k_1}{2}}$; $\forall s$

② $P(|X|>t) \le 2e^{-\frac{t^2}{2k_2}}$; $\forall t$

③ $E[e^{\lambda X^2}] \le 2$; $|\lambda| \le \frac{1}{k_3}$

k_1, k_2, k_3 مضاربی از یکدیگر هستند.

رسیدن از ① به ②: $Pr(|X|>t) \le e^{-\frac{t^2}{2k_1}}$ (در بالا ثابت شد) $\Rightarrow k_2 = k_1$

رسیدن از ② به ①: می‌توان نشان داد $c.k_1 \le k_2 \le k_1$

رسیدن از ③ به ②: چرنوف: $Pr(|X|>t) = Pr(X^2>t^2) = Pr(e^{\lambda X^2} > e^{\lambda t^2})$, $k_2 \le k_3$

رسیدن از ② به ③:

یادآوری: $EU = \int_0^\infty Pr(U>t)\,dt$; $U \geq 0$ اگر

اثبات: $\int_0^\infty Pr(U>t)\,dt = \int_0^\infty E[\mathbb{1}\{U>t\}]\,dt = E\int_0^\infty \mathbb{1}\{U>t\}\,dt = EU$

$$E[e^{\lambda X^2}] = \int_0^\infty Pr(e^{\lambda X^2}>t)\,dt = \int_0^\infty Pr\left(|X| > \sqrt{\frac{\ln t}{\lambda}}\right)dt$$

$$\overset{②}{\leq} \int_0^\infty \left(2e^{-\frac{\ln t}{\lambda . 2k_2}} \wedge 1\right)dt \quad ; \quad a \wedge b = \min(a,b)$$

$$= \int_0^\infty \left(2\left(\frac{1}{t}\right)^{\frac{1}{2\lambda k_2}} \wedge 1\right)dt \; ; \quad \lambda \le \frac{1}{k_3}$$

$$\leq \int_0^\infty \left(2\left(\frac{1}{t}\right)^{\frac{k_3}{2k_2}} \wedge 1\right) dt \quad ; \quad k_3 = 6k_2$$

$$= \int_0^\infty \left(2 \cdot \frac{1}{t^3} \wedge 1\right) dt = \int_0^1 dt + \int_1^\infty \frac{2}{t^3} dt = 2$$

* نرم اورلیز: Vershynin $\quad \|X\|_{\psi_2} = \min\left\{k_3 > 0 : E\left(e^{X^2/k_3^2}\right) \leq 2\right\}$

فرض کنید X، k_3ـ زیرگوسی (رابطه ③) باشد. همچنین Y، $\tilde{k}_3$ زیرگوسی باشد. در این صورت (تمرین): $\|X+Y\|_{\psi_2} \leq k_3 + \tilde{k}_3$

⊛ خاصیت زیرگوسی: فرض کنید $X_1, \dots, X_n$ متغیرهای تصادفی σ_i^2ـ زیرگوسی و مستقل باشند. آنگاه $S = \sum_{i=1}^n X_i$ نیز زیرگوسی با پارامتر $\sum_{i=1}^n \sigma_i^2$ خواهد بود.

(قضیه هوفدینگ)

حالت خاص: اگر $\sigma_i = \sigma$ باشد و لی X_i ها مستقل نباشند آنگاه S زیرگوسی با پارامتر حداکثر $n^2\sigma^2$ خواهد بود.

قضیه هوفدینگ در حالت استقلال: $\sigma_S^2 = n.\sigma^2$

اثبات: $E\left[e^{s(\sum X_i)}\right] = \prod E\left(e^{sX_i}\right) \leq \prod e^{\frac{s^2\sigma_i^2}{2}} = e^{\frac{s^2}{2}(\sum \sigma_i^2)}$

$$\Pr(|S| > t) \leq \exp\left(\frac{-t^2}{2\sum \sigma_i^2}\right)$$

* لم هوفدینگ: $X \in (a,b)$ ؛ X یک متغیر زیرگوسی با پارامتر $\frac{(b-a)^2}{4}$ است.

نکته: اگر $X \in (a,b)$، $Var(X) \leq \frac{(b-a)^2}{4}$.

اثبات: $Var(X) \leq E\left[(X-c)^2\right]$, $c = \frac{a+b}{2}$ قرار می‌دهیم

نتیجه لم هوفدینگ: $X_i \in (a_i, b_i)$ آنگاه (X_i ها مستقل)

$$\Pr\left[\left|\sum X_i - \mu_i\right| > t\right] \leq 2\exp\left(\frac{-2t^2}{\sum(b_i - a_i)^2}\right)$$

مثال: فرض کنید $X_i = \begin{cases} 1 & , p=\frac{1}{2} \\ -1 \end{cases}$. $\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$.

زیرگوسی با پارامتر n

$$\sum X_i \Rightarrow \Pr\left[\left| \frac{\sum X_i}{\sqrt{n}} \right| > t \right] \leq 2e^{-\frac{t^2}{2}}$$

جلسه هفتم: ۱۴۰۳, ۷, ۲۲

* اثبات لم هوفدینگ برای متغیرهای کران دار
* تعداد نمونه مورد نیاز برای تخمین میانگین
* متغیرهای زیرنمایی

لم هوفدینگ: $X \in [a,b]$ آنگاه X یک زیرگوسی با پارامتر $\frac{(b-a)^2}{4}$ است. $(O((b-a)^2))$

اثبات: (اثبات اول) زیرگوسی بودن با پارامتر ضعیف‌تر $(b-a)^2$.

$$E\left[e^{\lambda(X-EX)}\right] \leq e^{\lambda^2 \cdot \frac{\sigma^2}{2}}$$

ایده متقارن سازی: فرض کنید X و X' دو متغیر تصادفی مستقل و هم توزیع باشند.

$$E_X\left[e^{\lambda(X-E_{X'}X')}\right] \overset{\text{Jenson}}{\leq} E_X E_{X'}\left[e^{\lambda(X-X')}\right]$$

$$= E_X E_{X'} E_\xi\left[e^{\lambda\xi(X-X')}\right]$$

$$\xi = \begin{cases} 1 & ; \frac{1}{2} \\ -1 & ; \frac{1}{2} \end{cases} \Rightarrow \xi(X-X) \overset{d}{=} X-X'$$

فرض کنید $X=x$ ، $X'=x'$ باشد در این صورت:

$$= E_\xi\left[e^{\xi\lambda(x-x')}\right]$$

ادعا می‌کنیم: $\xi \sim Rad(\frac{1}{2})$ ، $\varphi_\xi(t) = E\left[e^{\xi t}\right] \leq e^{t^2/2}$.

$$\varphi_\xi(t) = \frac{e^t + e^{-t}}{2} \leq e^{t^2/2}$$ (تمرین بسط تیلور)

حال به ادامه اثبات باز می‌گردیم

$$E_\xi\left[e^{\xi\lambda(x-x')}\right] \leq e^{\frac{\lambda^2}{2}(x-x')^2} \leq e^{\frac{\lambda^2}{2}(b-a)^2}$$

چون $x - x' \leq b-a$

به اثبات دوم:

$$f(\lambda) = \log E\left[e^{\lambda(X-EX)}\right] \le \frac{\lambda^2\sigma^2}{2}$$

$$f(\lambda) = \overset{\circ}{\widehat{f(0)}} + \lambda f'(0) + \lambda^2 \frac{f''(z)}{2}, \quad z \in (0,\lambda)$$

$$f'(\lambda) = E\left[\frac{e^{\lambda(X-EX)}(X-EX)}{E[e^{\lambda(X-EX)}]}\right] \Rightarrow f'(0) = 0$$

$$f''(\lambda) = E\left[\frac{(X-EX)^2 e^{\lambda(X-EX)}}{E[e^{\lambda(X-EX)}]}\right] - \frac{\left[E\left(e^{\lambda(X-EX)}(X-EX)\right)\right]^2}{\left(E[e^{\lambda(X-EX)}]\right)^2}$$

فرض کنیم برای سادگی $EX=0$.

$p(x)$: f'' جمله اول $= \dfrac{\sum_x x^2 e^{\lambda x} p(x)}{\sum_x e^{\lambda x} p(x)} = \sum_x x^2 \dfrac{p(x) e^{\lambda x}}{\sum_{\bar{x}} p(\bar{x}) e^{\lambda \bar{x}}}$ $\quad p_\lambda(x)$: توزیع تیلت شده

$$= E_\lambda[X^2]$$

با ساده‌کردن جمله دوم می‌توان نوشت:

$$f''(\lambda) = E_\lambda[X^2] - (E_\lambda(X))^2 = Var_{p_\lambda}(X)$$

$$f(\lambda) = \frac{f''(z)}{2}\lambda^2 = \frac{Var_{p_z}[X]}{2}\lambda^2$$

$$\text{support}(P_z) = (a,b) \Rightarrow Var_{p_z}[X] \le \frac{(b-a)^2}{2} \Rightarrow f(\lambda) \le \frac{\lambda^2}{4}(b-a)^2$$

مثال: X زیرگوسی با پارامتر σ^2 و $EX = \mu$. فرض کنید μ را نمی‌دانیم. فرض کنید $X_1, \dots, X_n$ نمونه‌های X باشند.

$$\hat{\mu}(X_1, \dots, X_n) = \frac{\sum X_i}{n} \Rightarrow E[\hat{\mu}] = \mu$$

سؤال: چقدر نمونه لازم است تا با احتمال $1-\delta$ (مثلاً 95 درصد) تخمین تجربی در فاصله‌ی ϵ از میانگین واقعی قرار گیرد؟ $n(\epsilon, \delta)$

$$\frac{\sum X_i}{n} \sim \frac{\sigma^2}{n} \text{ subg}$$

$$\Pr\left[\left|\frac{\sum X_i}{n} - \mu\right| > \epsilon\right] \le \delta \Rightarrow e^{-\frac{\epsilon^2 \cdot n}{2\sigma^2}} \le \delta \Rightarrow n(\epsilon,\delta) \ge \frac{2\sigma^2}{\epsilon^2}\log \frac{1}{\delta}$$

اگر از نامساوی چبیشف استفاده می‌کردیم:

$$\frac{Var(X)}{n\epsilon^2} \le \frac{\sigma^2}{n\epsilon^2} < \delta \Rightarrow n \ge \frac{\sigma^2}{\epsilon^2}\frac{1}{\delta}$$

فرض کنید X یک متغیر با واریانس σ^2 باشد. بررسی مسئله پیچیدگی نمونه:

تخمین‌گر: میانه میانگین‌ها (قرار است پیچیدگی نمونه مشابه زیرگوسی داشته باشد)

← یک تخمین‌گر تجربی از میانگین داشته باشیم.

$$\underbrace{X_1,\dots,X_k}_{\hat{\mu}_1=\frac{\sum X_i}{k}},\ \underbrace{X_{k+1},\dots,X_{2k}}_{\hat{\mu}_2},\dots,\underbrace{X_{mk=n}}_{\hat{\mu}_m}$$

$$\hat{\mu} = \text{median}(\hat{\mu}_1,\dots,\hat{\mu}_m)$$

$$\Pr[|\hat{\mu}-\mu|>\epsilon] \le \Pr\left[\exists j_1,\dots,j_{\lceil \frac{m}{2}\rceil} : |\hat{\mu}_{j_i}-\mu|>\epsilon\right]$$

$$\le \sum_{\frac{m}{2}\text{ هر زیرمجموعه}} \Pr\left[\forall i=1,\dots,\tfrac{m}{2};\ |\hat{\mu}_{j_i}-\mu|>\epsilon\right]$$

$$= \binom{m}{\frac{m}{2}}\Big(\underbrace{\Pr[|\hat{\mu}_{j_1}-\mu|>\epsilon]}_{\frac{\sigma^2}{\epsilon^2 k}}\Big)^{\frac{m}{2}} \le 2^m\left(\frac{\sigma^2}{k\epsilon^2}\right)^{\frac{m}{2}} = \left(\frac{4\sigma^2}{k\epsilon^2}\right)^{\frac{m}{2}}$$

اگر قرار دهیم $k=\frac{16\sigma^2}{\epsilon^2}$ آنگاه

$$\Pr[|\hat{\mu}-\mu|>\epsilon] \le 2^{-m} < \delta \Rightarrow m > \log_2 \frac{1}{\delta}$$

$$n = km = O\left(\frac{\sigma^2}{\epsilon^2}\log_2\frac{1}{\delta}\right)$$

جلسه هشتم ۱۴۰۳, ۷, ۲۹

← متغیرهای زیرنمایی

← تمرکز اندازه برای χ^2 با d بعد

یادآوری: زیرگوسی

$$E[e^{\lambda(X-EX)}] \le e^{\frac{\lambda^2\sigma^2}{2}};\ \forall\lambda$$

فرض کنید $X \sim Exp(\beta)$. یعنی $f(x) = e^{-\beta x}\beta$ برای $x \geq 0$.

$$E[e^{\lambda X}] = \begin{cases} \dfrac{\beta}{\beta-\lambda} & ; \lambda < \beta \\ \infty & ; \lambda \geq \beta \end{cases}$$

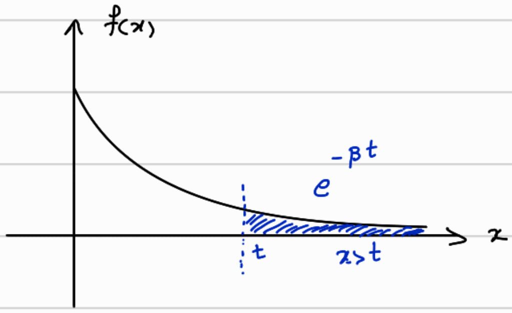
تعریف زیرنمایی: X یک متغیر (ν^2, α)-زیرنمایی است (از کتاب ون رایت) اگر

$$E[e^{\lambda(X-EX)}] \leq e^{\frac{\lambda^2 \nu^2}{2}} \; ; \; \forall |\lambda| < \frac{1}{\alpha}$$

نکته:

$$\log E[e^{\lambda(X-EX)}] \approx \frac{\lambda^2}{2} Var(X) + \dots$$

متغیر نمایی: فرض کنید $X \sim Exp(\beta)$

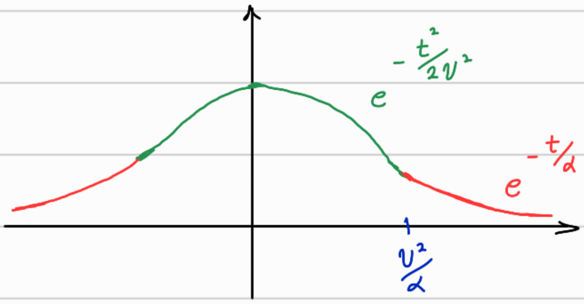


فرض کنید X یک متغیر (ν^2, α)-زیرنمایی باشد.

$$Pr[X-EX>t] \leq e^{-\lambda t} E[e^{\lambda(X-EX)}] \leq e^{-\lambda t + \frac{\lambda^2 \nu^2}{2}} \; ; \; \forall\, 0 \leq \lambda \leq \frac{1}{\alpha}$$

$$\overset{\min_{\lambda < 1/\alpha}}{=} \begin{cases} e^{-\frac{t^2}{2\nu^2}} & ; t \leq \frac{\nu^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & ; t > \frac{\nu^2}{\alpha} \end{cases}$$

اگر $EX = 0$ باشد باند بالا به شکل مقابل می‌شود:



* قضیه هوفدینگ زیرنمایی: فرض کنید $X_1, \dots, X_n$ هرکدام (ν_i^2, α_i)-زیرنمایی و مستقل باشند. فرض کنید $S = \sum X_i$ آنگاه S زیرنمایی با پارامترهای $(\sum \nu_i^2, \max \alpha_i)$ خواهد بود.

* مثال: $X \sim N(0,1)$. می‌خواهیم $Y = X^2$ را بررسی کنیم.

$$E[e^{\lambda(Y-\overset{1}{\overbrace{EY}})}] = e^{-\lambda} \overbrace{E[e^{\lambda x^2}]}^{\int e^{\lambda x^2} \cdot \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx} = \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \; ; \; \lambda < \frac{1}{2}$$

$$\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \overset{\text{بسط تیلور}}{\approx} e^{-\lambda}(1+\lambda+c\lambda^2) \approx e^{-\lambda} \cdot e^{\lambda+c\lambda^2} \Rightarrow \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \geq e^{c\lambda^2} \; \times \rightarrow$$ نامساوی در جهت برعکس

اگر c را 2 برابر کنیم نامساوی در جهت درست بدست می‌آید

$$\Rightarrow \frac{e^{-\lambda}}{\sqrt{1-2\lambda}} \leq e^{2\lambda^2} \; ; \; \lambda \leq \frac{1}{4} \Rightarrow$$ زیرنمایی $(4,4)$

* مثال: $\vec{Z} = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ و $X_i \sim N(0,1)$. فرض کنید X_iها مستقل نیز هستند.

$$\|Z\|^2 = \sum X_i^2 \; ; \; \chi_d^2 \overset{\text{قضیه هوفدینگ}}{\Rightarrow} \|Z\|^2 = \text{subE}(4d, 4)$$

$$E\|Z\|^2 = d \Rightarrow \frac{1}{d}\|Z\|^2 \sim \text{subE}\left(\frac{4}{d}, \frac{4}{d}\right)$$

$$\Pr\left[\left|\frac{1}{d}\|Z\|^2 - 1\right| \geq t\right] \leq 2 \begin{cases} e^{-\frac{dt^2}{8}} & t<1 \\ e^{-\frac{dt}{8}} & t>1 \end{cases} = 2e^{-\frac{d}{8}(t \wedge t^2)}$$

$$\Rightarrow \frac{1}{d}\|Z\|^2 \approx 1 + \frac{c}{\sqrt{d}} \Rightarrow \Pr(\dots) \leq 2e^{-\frac{c^2}{8}}$$

* کاربرد مهم: قضیه J-L برای کاهش بعد

10^6 نقطه $\rightarrow$ می‌توانند در $10^6 - 1$ بعد قرار بگیرند

$$\underset{\substack{\text{نگاشت}\\\text{کاهش بعد}}}{\mathcal{T}} : \mathbb{R}^D \rightarrow \mathbb{R}^d$$

می‌گوییم $\mathcal{T}$ یک نگاشت ϵ-حافظ فاصله است اگر

$$\mathcal{A} = \{P_1, \dots, P_n\}, \; \forall P_i, P_j : \frac{\|\mathcal{T}(P_i) - \mathcal{T}(P_j)\|}{\|P_i - P_j\|} \approx 1 \pm \epsilon$$

جانسون - لیندستراوس

* قضیه J-L: اگر $d > \frac{c}{\epsilon^2}\log n$ آنگاه یک نگاشت ϵ-حافظ فاصله وجود دارد.

اثبات: $\mathcal{T}$ نگاشت خطی

$$\frac{\|\mathcal{T}(P_i - P_j)\|}{\|P_i - P_j\|} \approx 1 \pm \epsilon$$

جلسه نهم ۱، 8، 1403

اثبات tight بودن قضیه J-L : Jelani Nelson (ترکیبی)

اثبات قضیه J-L : ایده 1 : استفاده از نگاشت خطی تصادفی $x \in \mathbb{R}^D : \underset{d\times D}{T}x$ ؛ $T = ($درایه‌های تصادفی$)$

هدف : $\Pr\left[\forall i,j : 1-\epsilon \leq \frac{\|T(P_i - P_j)\|}{\|P_i - P_j\|} \leq 1+\epsilon\right] > 0$ یا $1-\delta$

ماتریسی می‌خواهیم که نرم را حفظ کند: $E\,\|Tx\|^2 = \|x\|^2$

فرض کنید درایه‌های مستقل وجود داشته باشد: $T = \begin{pmatrix} & t_{ij} & \\ & & \end{pmatrix}$

$T_{ij} \sim \mathcal{N}(0, \sigma^2)$

$$\|Tx\|^2 = \left\|\begin{pmatrix} T_1^T \\ \vdots \\ T_d^T \end{pmatrix} x\right\|^2 = \left\|\begin{pmatrix} \langle T_1, x\rangle \\ \vdots \\ \langle T_d, x\rangle \end{pmatrix}\right\|^2 = \sum_{j=1}^{d} \langle T_j, x\rangle^2$$

$$\langle T_j, x\rangle = \sum t_{ij} x_i \sim \mathcal{N}(0, \sigma^2 \|x\|^2)$$

$$E \sum_{j=1}^{d} \langle T_j, x\rangle^2 = d\,\sigma^2 \|x\|^2 \overset{\text{برای حفظ شدن اندازه}}{\Longrightarrow} \sigma^2 = \frac{1}{d}, \quad \langle T_j, x\rangle \sim \mathcal{N}\left(0, \frac{\|x\|^2}{d}\right)$$

$$A := \left\{\frac{P_i - P_j}{\|P_i - P_j\|}, \forall i,j\right\}, \quad |A| = \binom{n}{2}$$

هدف دیگر: $\Pr\left[\forall a \in A : 1-\epsilon < \|Ta\| < 1+\epsilon\right] > 1-\delta$ یا $p = \Pr\left[\exists a \in A : \left|\|Ta\|^2 - 1\right| > \tilde{\epsilon}\right] < \delta$

$$p < \sum_{a \in A} \Pr\left[\left|\|Ta\|^2 - 1\right| > \tilde{\epsilon}\right]$$

$$\|Ta\|^2 = \sum_{j=1}^{d} \langle T_j, a\rangle^2 = \text{جمع } d \text{ تا مجذور مستقل } \mathcal{N}(0, \tfrac{1}{d}) = \frac{1}{d}\left(\text{جمع } d \text{ تا مجذور نرمال } \mathcal{N}(0,1)\right)$$

جلسه قبل : $Z_1, \ldots, Z_d \sim \mathcal{N}(0,1) \Longrightarrow \Pr\left[\left|\frac{1}{d}\|Z\|^2 - 1\right| > \epsilon\right] < e^{-\frac{d}{8}(\epsilon \wedge \epsilon^2)}$

$$\Rightarrow \quad p < n^2 \cdot e^{-\frac{d}{8}\epsilon^2} \qquad (\text{با فرض } \epsilon < 1)$$

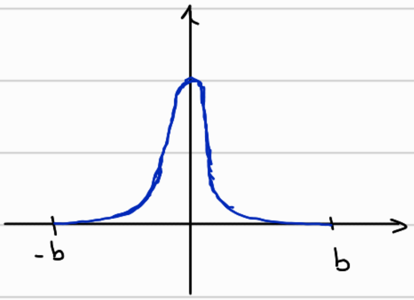
$$p < \delta \Rightarrow \quad n^2 e^{-\frac{d}{8}\epsilon^2} < \delta \Rightarrow \quad d \gtrsim \frac{1}{\epsilon^2} \log\left(\frac{n}{\sqrt{\delta}}\right)$$

* نامساوی برنشتاین
* نامساوی McDiarmid

آیا متغیرهای تصادفی کران دار، زیرنمایی هستند؟

دقت کم هوفدینگ برای متغیرهای کران دار، چطور خوب است؟

فرض کنید $|X| < b$ پس X زیرگوسی با پارامتر b^2 است.



$\sigma^2 << b^2$

$\rightarrow$ هوفدینگ باند خوبی نمی دهد

* هر متغیر کران دار با واریانس σ^2 و عرض b زیرنمایی خواهد شد.

با احتمال 1: $|X - EX| < b$

$\mathrm{Var}\, X = \sigma^2$

هدف: $\log E\left[e^{\lambda(X-EX)}\right] < \frac{\lambda^2 \nu^2}{2}, \quad |\lambda| < \frac{1}{\alpha}$

$$E\left[e^{\lambda(X-EX)}\right] = 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} E\left[(X-EX)^k\right]$$

$$\leq 1 + \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} E\big[\underbrace{|X-EX|^k}_{\leq b^{k-2} \cdot |X-EX|^2}\big]$$

$$= 1 + \sigma^2 \sum_{k=2}^{\infty} \frac{\lambda^k}{k!} b^{k-2} \underset{k>2}{\leq} 1 + (\lambda\sigma)^2 \sum_{k=2}^{\infty} \frac{(|\lambda| b)^{k-2}}{2}$$

$$= 1 + \frac{\lambda^2\sigma^2}{2(1-|\lambda| b)}, \quad |\lambda| < \frac{1}{b}$$

$$< \exp\left(\frac{\lambda^2\sigma^2}{2(1-|\lambda| b)}\right), \quad |\lambda| < \frac{1}{b}$$

$$< \exp(\lambda^2\sigma^2) \; ; \; |\lambda| < \frac{1}{2b}$$

$$\Rightarrow \quad X \sim s\text{-}Exp(2\sigma^2, 2b) \Rightarrow \Pr(|X - EX| > t) = \begin{cases} e^{-\frac{t^2}{4\sigma^2}}, & t < \frac{\sigma^2}{b} \\ e^{-\frac{t}{4b}}, & \frac{\sigma^2}{b} < t < b \end{cases}$$

نامساوی بنت (ون رایت) : متغیرهای کران دار، با ممان دوم محدود زیرپواسون هستند.

* نامساوی برای توابع دلخواه از n متغیر تصادفی با خاصیت تفاضل محدود. $f(x_1, \dots, x_n)$

$$\forall x_1, x_1', x_{2:n} : f(x_1, \dots, x_n) - f(x_1', x_2, \dots, x_n) \leq c_1$$

$$\forall i, x_i, x_i', x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n : f(x_{-i}, x_i) - f(x_{-i}, x_i') \leq c_i$$

قضیه McDiarmid (Azuma - Hoeffding) : اگر f تفاضل محدود با $(c_1, \dots, c_n)$، $X_1, \dots, X_n$ مستقل باشند آنگاه متغیر

$Z = f(X_1, \dots, X_n)$ زیرگوسی با پارامتر $\sum c_i^2$ خواهد بود.

مثال: $S = \frac{\sum X_i}{n}$; $X_i \in (0,1)$ تفاضل محدود با پارامتر $(\frac{1}{n}, \dots, \frac{1}{n})$ است.

جلسه دهم 1403,8,6

اثبات قضیه McDiarmid : با استقرا روی تعداد متغیرهای تصادفی

$n = 1$: $Z = f(X_1)$. اگر f، c_1 تفاضل محدود باشد، f، c_1-زیرگوسی است. f داخل بازه به طول c_1، $\min f \leq f \leq \max f$

$\Rightarrow$ طبق لم هوفدینگ ✓

حال فرض کنید حکم برای $n-1$ درست باشد. $\leftarrow$ $S \leq n$

$$\underset{X_1 \dots X_n}{E}\left[e^{\lambda(f(X_{1:n}) - Ef(X_{1:n}))} \right] \leq e^{\frac{\sum c_i^2}{8}\lambda^2}$$

$$= E_{X_1 \dots X_{n-1}} E_{X_n} \left[e^{\lambda [f(X_{1:n}) - E_{X_n} f(X_{1:n}) + \underbrace{E_{X_n} f(X_{1:n})}_{h(X_{1:n-1})} - E f(X_{1:n})]} \right]$$

$$= E_{X_{1:n-1}} \left[e^{\lambda (h(X_{1:n-1}) - E f(X_{1:n}))} \underbrace{E_{X_n} \left(e^{\lambda (f(X_{1:n}) - E_{X_n} f(X_{1:n}))} \right)}_{\leq e^{\frac{\lambda^2 c_n^2}{8}}} \right]$$

$E[g(X,Y)] = E_X [E_{Y|X} (g(X,Y))]$

$$= e^{\frac{\lambda^2 c_n^2}{8}} \underbrace{E_{X_{1:n-1}} \left[e^{\lambda (h(X_{1:n-1}) - E h)} \right]}_{\leq e^{\frac{\lambda^2}{8} \sum_{i=1}^{n-1} c_i^2}}$$

اگر h تفاضل محدود

$c_1, \dots, c_{n-1}$

را از f به ارث ببرد طبق فرض

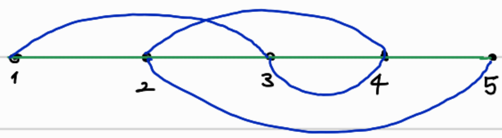
استقرا

$$h(x_1, \dots, x_{n-1}) - h(x_1', \dots, x_{n-1}) = \sum_{x_n} p(x_n) f(x_1, \dots, x_n) - \sum_{x_n} p(x_n) f(x_1', \dots, x_n)$$

$$= \sum p(x_n) \overbrace{(f(x_1, \dots) - f(x_1', \dots))}^{c_1} \leq c_1$$

مثال : فرض کنید گراف تصادفی $G(n,p)$ داشته باشیم. تعداد مسیرهای خزانگر را (L) بیابید. مسیر خزانگر یعنی مسیری که تمام رأس ها داخل آن قرار داشته باشند.

مثال از مسیرها :



$$E L = E \sum_{\substack{\text{تمام جایگشت ها} \\ (i_1, \dots, i_n)}} 1 \{ i_1 \to i_2 \to \dots \to i_n \} = n! \, p^{n-1} \approx \left(\frac{np}{e} \right)^n \to \quad p > \frac{e}{n} \text{ یا } np > e$$

اگر

باشد میانگین نامحدود است

$$L = L(\overset{\text{یال ها}}{e_{12}}, \dots, e_{1n}, e_{23}, \dots, e_{2n}, \dots, e_{n-1,n})$$

آیا می توان در مورد تمرکز L روی میانگین صحبت کرد ؟

تعداد مسیرهایی که رأس های 1 و 2 را با هم بگیریم

$$L(e_{12}, \dots, e_{n-1,n}) - L(e_{12}', \dots, e_{n-1,n}) \leq 2 \overbrace{(n-1)!} \xrightarrow{\text{زیرگوسی}} \sum c_i^2 = \binom{n}{2} (2(n-1)!)^2$$

حال تعداد مثلث های گراف را محاسبه کنیم :

$$E \Delta = \binom{n}{3} p^3 \approx n^3 p^3$$

$$\Delta(e_{12},\dots)-\Delta(e'_{12},\dots)\le n-2\approx n \xrightarrow{\text{از گوس}} \frac{\sum c_i^2}{4}=\binom{n}{2}\cdot\frac{n^2}{4}<\frac{n^4}{8}$$

$$P(|\Delta-E\Delta|>t)\le 2e^{-\frac{4t^2}{n^4}};$$ اگر $t=O(n^2)$ باشد احتمال ناچیز می‌شود

پس برای pهای غیرناچیز مانند $p>0.1$، $\Delta=n^3\pm O(n^2)$

مثال: تخمین آنتروپی

یک متغیر تصادفی $X\in\{1,\dots,k\}$ درنظر بگیرید.

$$X\sim p_X:\quad H(P_X)=-\sum_{j=1}^{k}P(X=j)\log P(X=j)$$

فرض کنید $X_1,X_2,\dots,X_n$ نمونه‌های توزیع P_X باشند. می‌خواهیم آنتروپی را تخمین بزنیم.

$$\hat{H}(X_1,X_2,\dots,X_n)=H(P_{emp}(X_1,\dots,X_n))$$

$$P_{emp}(X=a)=\frac{1}{n}\sum_{i=1}^{n}1\{X_i=a\}$$

سؤال 1: $E\hat{H}$ با H چه رابطه‌ای دارد؟ سؤال 2: در مورد تفاضل محدود بودن $\hat{H}$ چه می‌توان گفت؟

$$E\hat{H}=E\Big[-\sum_{a=1}^{k}\overbrace{P_{emp}(a)\log P_{emp}(a)}^{\text{تابع محدب نسبت به } P_{emp}}\Big]$$

$$\overset{\text{Jenson}}{\le}-\sum\underbrace{EP_{emp}(a)}_{P(a)}\log\underbrace{EP_{emp}(a)}_{P(a)}=H\Rightarrow\quad 0\le H-E\hat{H}\le\overset{\text{تمرین}}{\frac{k}{n}}$$

تفاضل محدود بودن $\hat{H}$:

$$\hat{H}_1(x_1=a,x_2,\dots,x_n)-\hat{H}_2(x_1=b,x_2,\dots,x_n)$$

چون تغییر x_1 فقط توزیع تجربی a و b را عوض می‌کند

$$=-\hat{P}_1(a)\log\hat{P}_1(a)-\hat{P}_1(b)\log\hat{P}_1(b)-\cancel{\sum_{j\ne a,b}\hat{P}_1(j)\log\hat{P}_1(j)}-\Big(-\hat{P}_2(a)\log\hat{P}_2(a)-\hat{P}_2(b)\log\hat{P}_2(b)-\cancel{\sum_{j\ne a,b}\hat{P}_2(j)\log\hat{P}_2(j)}$$

$$\begin{cases}\hat{P}_1(a)-\hat{P}_2(a)\le\frac{1}{n}\\ \hat{P}_2(b)-\hat{P}_1(b)\le\frac{1}{n}\end{cases}$$

تمرین: اگر $x-y \le t$ باشد آنگاه $|x\log x - y\log y| \le \max(-t\log t, -(1-t)\log(1-t))$

چون در تابع ما $t=\frac{1}{n}$ پس خواهیم داشت: $c_i = \frac{\log n}{n} \Rightarrow$ پارامتر زیرگوسی $= \frac{\sum c_i^2}{4} = \frac{(\log n)^2}{n}$

انحراف معیار $\le \frac{\log n}{\sqrt{n}} \Rightarrow H \approx E\hat{H} + \frac{k}{n} + c.\frac{\log n}{\sqrt{n}}$ با احتمال بالا

* جلسه یازدهم ۸، ۸، ۱۴۰۳

واریانس

* کران بالا
* مثال
* نامساوی پوانکاره

* فرض کنید $Z = f(X_1,\dots,X_n)$

$$Var(Z) \le E_{X_1,\dots,X_n} \sum_{i=1}^{n} \underbrace{Var_{X_i} f(X_1,\dots,X_n)}^{g(X_{-i})}$$

اتحاد: $Var(X) = \frac{1}{2}E[(X-X')^2]$; X, X' هم‌توزیع

اتحاد: $Var(X) = E[((X-X')_+)^2]$; $(t)_+ = relu(t)$

فرم‌های دیگر نامساوی پوانکاره:

$$Var\, f \le E_{\substack{X_1,\dots,X_n \\ X'_1,\dots,X'_n}} \sum_{i=1}^{n} (f(X_1,\dots,X_n) - f(X_{-i}, X'_i))_+^2$$

هم توزیع بودن

اثبات استقرایی:

برای $n=2$:

$$E_{X_1X_2}\left[(f(X_1,X_2) - E_{X_1}f(X_1,X_2) - E_{X_2}f(X_1,X_2) + E_{X_1X_2}f(X_1,X_2))^2\right] \ge 0$$

$$E_{X_2}Var_{X_1}f + E_{X_1}Var_{X_2}f - Var f \ge 0 \checkmark$$

حال به سراغ گام استقرایی رویم. فرض کنید حکم برای هر تابعی از $n-1$ متغیر مستقل درست باشد.

مرحله اول

$X_n \,,\, (X_1, \dots, X_{n-1})$

$$\operatorname{Var} f \leq \mathbb{E}_{X_n} \underbrace{\operatorname{Var}_{X_1 \dots X_{n-1}}} f + \mathbb{E}_{X_1 \dots X_{n-1}} \operatorname{Var}_{X_n} f$$

$$\operatorname{Var}_{X_1 \dots X_{n-1}} f(X_1, \dots, X_{n-1}, x_n) \leq \sum_{i=1}^{n-1} \mathbb{E} \operatorname{Var}_{X_i} f(X_1, \dots, X_{n-1}, x_n)$$

فرض استقرا

مثال: قضیه پوانکاره: $\vec{X} \sim \mathcal{N}(0, I_d)$; (X_iها d بعدی هستند)

$$\operatorname{Var} f(\vec{X}) \leq \mathbb{E}_{\vec{X}} \left[\| \nabla f(\vec{X}) \|^2 \right]$$

حالت خاص: اگر f، L-لیپ شیتز باشد $\| \nabla f \| \leq L$ (برای f مشتق پذیر)

$$|f(x) - f(y)| \leq L \cdot \| x - y \|$$

در این حالت می‌توان گفت $\operatorname{Var} f \leq L^2$

اثبات قضیه پوانکاره:

$$\operatorname{Var} f(X_1, \dots, X_d) \leq \sum_{i=1}^{d} \mathbb{E} \left(\frac{\partial f}{\partial x_i} (X_1, \dots, X_d) \right)^2$$

فرض کنید برای تابع یک متغیره حکم درست باشد. یعنی $\operatorname{Var} f(x) \leq \mathbb{E} f'(x)^2$

$$\xrightarrow{\text{نامساوی واریانس}} \operatorname{Var} f \leq \mathbb{E}_X \sum \operatorname{Var}_{X_i} f(X_i, X_{-i}) \overset{\text{حالت } d=1}{\leq} \mathbb{E}_{X_{-i}} \sum \mathbb{E}_{X_i} \left(\frac{\partial}{\partial x_i} f(X_i, X_{-i}) \right)^2$$

حال حالت یک بعدی را ثابت می‌کنیم: $X \sim \mathcal{N}(0,1) \,,\, f(x)$

CLT: $X \overset{dist}{\underset{n \to \infty}{\simeq}} \frac{1}{\sqrt{n}} (\xi_1, \dots, \xi_n)$; $\xi_i = \pm 1$ با احتمال مساوی

$$\operatorname{Var} f(X) \approx \operatorname{Var}\left(\frac{1}{\sqrt{n}} (\xi_1, \dots, \xi_n) \right) \leq \mathbb{E}_{\vec{\xi}} \sum \operatorname{Var}_{\xi_i} f\left(\frac{1}{\sqrt{n}} (\xi_1 + \dots + \xi_n) \right)$$

$$= \mathbb{E}_{\vec{\xi}} \sum \frac{1}{4} \left[f\left(\frac{\sum_{j \neq i} \xi_j + 1}{\sqrt{n}} \right) - f\left(\frac{\sum_{j \neq i} \xi_j - 1}{\sqrt{n}} \right) \right]^2$$

$$\approx \mathbb{E}_{\vec{\xi}} \sum_{i=1}^{n} \frac{1}{4} \left(f'\left(\frac{\sum_{j \neq i} \xi_j}{\sqrt{n}} \right) \right)^2 \cdot \frac{4}{n}$$

$$\approx \mathbb{E}_{\vec{\xi}} f'\Big(\underbrace{\frac{\sum \xi_i}{\sqrt{n}}}_{\approx X} \Big)^2 = \mathbb{E} f'(X)^2$$

مثال: بزرگترین مقدار ویژه‌ی یک ماتریس تصادفی متقارن

$$A = \underbrace{\begin{pmatrix} \pm 1 & \pm 1 & \\ \pm 1 & \ddots & \\ & & \end{pmatrix}}_{n \text{ در } n} \qquad \pm 1 \text{ با احتمال } \tfrac{1}{2} \text{ و مستقل}$$

$E\,\lambda_{max}(A) = \Theta(\sqrt{n})$: جلسات آینده

می‌خواهیم درمورد واریانس آن صحبت کنیم. $Var\,\lambda_{max}(A) \leq C$

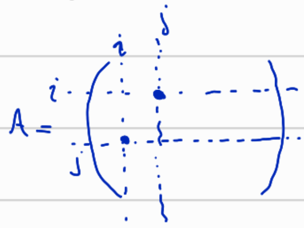
$$\lambda_{max}(A) = \sup_{u:\, \|u\|=1} u^T A u$$

$$\lambda_{max} = \lambda_{max}(A_{11}, A_{12}, \dots, A_{1n}, A_{22}, A_{23}, \dots, A_{2n}, \dots, A_{nn})$$

$\frac{n(n+1)}{2}$ متغیر

± 1 مستقل

$$Var\,\lambda_{max} \leq \sum_{1\leq i\leq j\leq n} E\, Var_{A_{ij}}\, \lambda_{max}(A)$$



$$= \sum \underset{\substack{A_{11},\dots,A_{nn}\\ A'_{11},\dots,A'_{nn}}}{E} \left[\left(\lambda_{max}(A) - \lambda_{max}(A_{\sim ij}, A'_{ij}) \right)_+^2 \right]$$

$A^{(ij)} \rightarrow$ در درایه ij متفاوت است

$$= \sum E\left[\left(\sup u^T A u - \sup u^T A^{(ij)} u \right)_+^2 \right], \quad u^* = \arg\max\, u^T A u$$

$$\leq \sum_{1\leq i\leq j\leq n} E\left[\left(u^{*T} A u^* - u^{*T} A^{ij} u^* \right)_+^2 \right], \quad u^{*T}(A - A^{ij})u^* = \sum u_k^* u_l^* (A - A^{ij})_{kl}$$

$$\leq E \sum_{1\leq i\leq j\leq n} 16\, u_i^{*2} u_j^{*2}$$

$$(u^{*T}(A-A^{ij})u^*)^2 \leq \begin{cases} 4u_i^{*4} & ;\ i=j \\ 16\, u_i^{*2} u_j^{*2} & ;\ i\neq j \end{cases} \leq 16\, u_i^{*2} u_j^{*2}$$

$$\leq 16\, E \underbrace{\sum_{i,j} u_i^{*2} u_j^{*2}}_{(\sum u_i^{*2})(\sum u_j^{*2})} = 16$$

جلسه دوازدهم 13_8_1403

- زیرگوسی بودن توابع لیپشیتز از نرمال
- محاسبه امیدریاضی سوپریمم

قضیه پوانکاره: $f: \mathbb{R}^d \rightarrow \mathbb{R}$ ، $X \sim N(0, I_d)$:

$$Var(f(X)) \leq \underset{X\sim N(0,I)}{E}\left[\|\nabla f(X)\|_2^2 \right]$$

حالت خاص: اگر f تابع L_لیپشیتز باشد آنگاه $Var(f(X)) \leq L^2$.

ـ قضیه: اگر $X\sim N(0,I_d)$ و f یک تابع L-Lip باشد آنگاه، $Z=f(x)$ ، L^2 زیرگوسی می‌شود.

$$\Pr[\,|f(x)-Ef(x)|>t\,] \le 2e^{-\frac{t^2}{2L^2}}$$

مثال: $f(\vec{x})=\max_{1\le i\le d} X_i$ ، $\mathrm{Var}\, f(x)$??

$\ell=\arg\max x_i$

$$f(\vec{x})-f(\vec{y})=\max_{1\le i\le d} x_i-\max_{1\le i\le d} y_i = x_\ell-\max_{1\le i\le d} y_i$$

$$\le x_\ell-y_\ell \le \|\vec{x}-\vec{y}\|_2 \Rightarrow$$ f تابع 1-Lip است

$$\Rightarrow \mathrm{Var}(f(x))\le 1,\quad \Pr[\max X_i - E\max X_i \ge t]\le e^{-\frac{t^2}{2}}$$

مثال: $Z=\frac{1}{d}\sum_{i=1}^{d} X_i^2$ ، $X\sim N(0,I_d)$. قبلاً دیدیم $\Pr[\,|Z-1|>t\,]\le 2e^{-\frac{d}{8}\min(t,t^2)}$.

$Y=\sqrt{\sum_{i=1}^{d} X_i^2}=\|\vec{X}\| \longrightarrow$ طبق تعریف 1-Lip $\Rightarrow$ 1 زیرگوسی نیز هست $\Rightarrow \Pr[Y-EY>t]\le e^{-\frac{t^2}{2}}$

$$EY\le\sqrt{EY^2}=\sqrt{d}\Rightarrow \Pr(Y-\sqrt{d}>t)\le e^{-\frac{t^2}{2}} \Rightarrow \Pr\Big(\overbrace{\frac{1}{\sqrt{d}}Y}^{\sqrt{Z}}-1>\overbrace{\frac{t}{\sqrt{d}}}^{t'}\Big)\le e^{-\frac{t^2}{2}}$$

$$\Rightarrow \Pr\big(Z>\underbrace{(1+t')^2}_{\le 1+3t'}\big)\le e^{-\frac{d}{2}t'^2}$$

اگر $t'<1$

$\tilde{t}=3t'$

$$\Rightarrow \Pr(Z>1+\tilde{t})\le e^{-\frac{d\tilde{t}^2}{18}}\ ;\quad \tilde{t}<3$$

ـ محاسبه‌ی امید ریاضی سوپریمم از متغیرهای تصادفی

ـ $\{X_t\}_{t\in T}$ ، $E[\sup_{t\in T} X_t]$ (مثلاً $T=\mathbb{R}$)

ـ ابتدا فرض کنید T یک مجموعه‌ی متناهی باشد. هر مجموعه‌ی نامتناهی را نیز می‌توان با تقسیم‌بندی به یک مجموعه‌ی متناهی تبدیل کرد.

ـ هدف: $X_1,\dots,X_n$ متغیر تصادفی (نه لزوماً مستقل)

$$E[\max_{1\le i\le n} X_i]\le E[\sum_{i=1}^{n} X_i]=\sum_{i=1}^{n} E(X_i)$$

اگر X_iها iid باشند: $E[\max_{1\le i\le n} X_i]\le nE(X)\le E(\sqrt{\sum X_i^2})\le\sqrt{E\sum X_i^2}=\sqrt{nEX^2}$

اگر $p>1$: $\max X_i \le \underbrace{\left(\sum X_i^p\right)^{1/p}}_{\text{مقعر}} \Rightarrow E[\max X_i] \le n^{1/p} \underbrace{\left(E X^p\right)^{1/p}}_{\|X\|_{L^p}}$

روش دیگر باند زدن :

$$\max X_i = \frac{1}{\beta} \log \max e^{\beta X_i}$$

$$\le \frac{1}{\beta} \log \sum e^{\beta X_i} \Rightarrow E(\max_{1\le i\le n} X_i) \le \frac{1}{\beta} E\left[\log \sum e^{\beta X_i}\right]$$

$$\Rightarrow \boxed{E(\max_{1\le i\le n} X_i) \le \frac{1}{\beta} \log\left(\sum E[e^{\beta X_i}]\right)}$$

فرض کنید X_i میانگین صفر و σ^2 زیرگوسی باشد.

$$E[e^{\beta X_i}] \le e^{\beta^2 \frac{\sigma^2}{2}} \Rightarrow E[\max X_i] \le \frac{1}{\beta} \log\left(n e^{\beta^2 \frac{\sigma^2}{2}}\right)$$

$$= \frac{1}{\beta} \log n + \frac{\beta \sigma^2}{2} \le \boxed{\sqrt{2\sigma^2 \log n}}$$

کران مقابل زمانی که X_i ها مستقل گوسی باشند tight است

مثال، اگر $X_i \sim N(0,1)$ باشد، $E[\max X_i] \approx \sqrt{2\log n}$ و دیدیم که $Var(\max X_i) \le 1$ پس $\max X_i$ خیلی نزدیک به میانگین است.

جلسه سیزدهم 15/8/1403

- امید ریاضی سوپریمم
- فرآیند لیپشیتز
- کران بالا روی فرآیند Lip
- مثال: نرم ماتریس تصادفی

- فرآیند لیپ شیتز : $\{X_t \in \mathbb{R}\}_{t\in T}$ ؛ فرض کنید فضای T مجهز به متر d باشد. (مثلاً $\mathbb{R}^2$ با متر اقلیدسی) فرآیند را لیپ شیتز می‌گوییم اگر یک متغیر تصادفی C وجود داشته باشد که $|X_t - X_s| \le C \cdot d(s,t)$.

بررسی C :

$X: \Omega \to \mathbb{R}$, $X(\omega)$ $\quad |X_t(\omega) - X_s(\omega)| \le C(\omega)\, d(s,t)$

مثال: فرض کنید A یک ماتریس با درایه‌های مستقل A_{ij} باشد به طوری که $A_{ij} \sim subg(\sigma^2)$. $\le E[\|A\|_{op}]$? $(A_{m\times n})$ فرض کنید میانگین A_{ij} صفر باشد.

$$Y_{u,v} \;;\; t=(u,v) \;,\; \mathcal{T} = B_m^{(1)} \times B_n^{(1)}$$

نقاط داخل کره واحد m بعدی

$$\|A\|_{op} = \sup_{\substack{u \in \text{کره واحد } m \text{ بعدی} \\ v \in \text{کره واحد } n \text{ بعدی}}} \overbrace{u^T A v}^{Y_{u,v}}$$

می‌خواهیم نشان دهیم این فرآیند Lip است.

$$\forall \begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde u \\ \tilde v \end{bmatrix} : \quad |Y_{u,v} - Y_{\tilde u,\tilde v}| \le C \cdot d\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde u \\ \tilde v \end{bmatrix}\right)$$

$$|Y_{u,v} - Y_{\tilde u,\tilde v}| = |u^T A v - \tilde u^T A \tilde v| \le |\overbrace{u^T A v - u^T A \tilde v}^{u^T A(v-\tilde v)}| + |\overbrace{u^T A \tilde v - \tilde u^T A \tilde v}^{(u-\tilde u)^T A v}|$$

$\|u\| \le 1$

یادآوری: $|x^T B y| \le \|x\| \cdot \|B\| \cdot \|y\|$

$$\le \|A\| \, \|v-\tilde v\| + \|u-\tilde u\| \, \|A\| = \underbrace{\|A\|}_{C} \Big(\underbrace{\|v-\tilde v\| + \|u-\tilde u\|}_{\triangleq \, d\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde u \\ \tilde v \end{bmatrix}\right)} \Big)$$

← حال فرض کنید $\{X_t\}$ فرآیند لیپ شیتز با متغیر تصادفی C باشد. همچنین X_tها هم زیرگوسی σ^2 با میانگین صفر باشند.

$$E\Big[\sup_{t \in \mathcal{T}} X_t\Big] \le \epsilon \cdot \mathbb{E}[C] + \sqrt{2\sigma^2 \log N(\mathcal{T}, d, \epsilon)}$$

← عدد پوشش: فضای $\mathcal{T}$ با متر d را درنظر بگیرید. مجموعه $\{Q_1, \dots, Q_L\}$ را یک ϵ-پوشش از $\mathcal{T}$ می‌گوییم اگر $\mathcal{T} \subseteq \bigcup_{i=1}^{L} B_d(Q_i ; \epsilon)$

کمترین تعداد نقاط لازم برای پوشش $\mathcal{T}$ در شعاع ϵ را ϵ-عدد پوششی می‌گوییم و با $N(\mathcal{T}, d, \epsilon)$ نمایش می‌دهیم.

اثبات: اگر $\{Q_1, \dots, Q_L\}$ یک ϵ-پوشش باشد نزدیک‌ترین Q_i به t را با $\pi(t)$ نشان می‌دهیم.

$$\sup_{t \in \mathcal{T}} X_t = \sup_{t \in \mathcal{T}} \left(X_t - X_{\pi(t)} + X_{\pi(t)} \right)$$

$$\le \sup_{t \in \mathcal{T}} \left(X_t - X_{\pi(t)} \right) + \sup_{t \in \mathcal{T}} \left(X_{\pi(t)} \right)$$

$$= \sup_{t \in \mathcal{T}} \left(X_t - X_{\pi(t)} \right) + \max_{\{Q_i\}} \left(X_{Q_i} \right)$$

$$E\Big[\sup_{t \in \mathcal{T}} X_t\Big] \le E\Big[\sup_{t \in \mathcal{T}} (\underbrace{X_t - X_{\pi(t)}}_{\le \epsilon C})\Big] + \sqrt{2\sigma^2 \log L} \le \epsilon \, E(C) + \sqrt{2\sigma^2 \log L}$$

حال دوباره به مسئله نرم ماتریس برمی‌گردیم. ابتدا بررسی می‌کنیم $Y_{u,v}$، σ^2-زیرگوسی هست یا خیر.

$$Y_{u,v} = u^T A v = \sum_{i,j} \overbrace{u_i v_j A_{ij}}^{\text{subg}(u_i^2 v_j^2 \sigma^2)}$$; A_{ij}ها مستقل و σ^2-زیرگوسی

$\|u\|^2 \|v\|^2 \leq 1$

قضیه هوفدینگ

$$= \text{subg}\left(\sigma^2 \sum_{i,j} u_i^2 v_j^2\right) = \text{subg}(\sigma^2)$$

طبق قضیه :

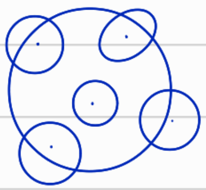
$$E\|A\| = E \sup Y_{u,v} \leq \epsilon . E\|A\| + \sqrt{2\sigma^2 \log N(B_m^{(1)} \times B_n^{(1)}, d_{\text{تعریف شده}}, \epsilon)}$$

$$\Rightarrow \quad E\|A\| \leq \frac{1}{1-\epsilon} \sqrt{2\sigma^2 \log N(B_m^{(1)} \times B_n^{(1)}, d, \epsilon)}$$

محاسبه عدد پوشش : فرض کنید کره‌ای اقلیدسی به شعاع 1 داریم. می‌خواهیم عدد پوشش آن را حساب کنیم.

$$\left(\frac{1}{\epsilon}\right)^m \leq N(B_m(1), \|\ \|_2, \epsilon) \leq \left(\frac{3}{\epsilon}\right)^m \quad (\epsilon<1)$$

روش بدست آوردن ، بازی با حجم

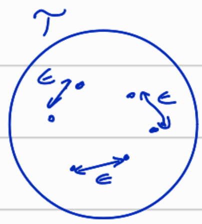


حجم کره واحد $\leq$ حجم کره‌های پوشش دهنده به شعاع ϵ

$$N . k . (\epsilon)^m \geq k . (1)^m \Rightarrow N \geq \left(\frac{1}{\epsilon}\right)^m$$

covering

عدد پوششی : مینیمم تعداد نقاط لازم برای ϵ-پوشش



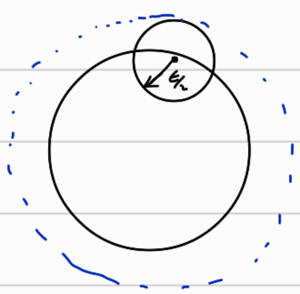
عدد گنجایش : حداکثر تعداد نقاط از T که فاصله دو به دوی آن‌ها حداقل ϵ باشد. $P(\epsilon)$

packing

* نامساوی : $N(\epsilon) \leq P(\epsilon)$

فرض کنید $\{Q_1, \dots, Q_L\}$ یک ϵ-گنجایش ماکسیمال باشد. آن وقت $\{Q_1, \dots, Q_L\}$ یک ϵ-پوشش هم هست. چون اگر نقطه‌ای در فاصله بیش از ϵ از همه نقاط Q وجود داشته باشد ماکسیمال بودن مجموعه نقض می‌شود.

حال عدد گنجایشی رو واحد را باند می‌زنیم.



$$\text{حجم کره } 1+\frac{\epsilon}{2} \geq \text{حجم کره به شعاع } \frac{\epsilon}{2} \times P$$

$$P \leq \left(\frac{1+\frac{\epsilon}{2}}{\frac{\epsilon}{2}}\right)^m = \left(1+\frac{2}{\epsilon}\right)^m \leq \left(\frac{3}{\epsilon}\right)^m$$

دوباره به مسئله نرم ماتریس برمی‌گردیم.

$$N\left(B_m^{(1)} \times B_n^{(1)}, \|u-\tilde{u}\| + \|v-\tilde{v}\|, \epsilon\right)$$

فرض کنید $\{u_1, \dots, u_L\}$ یک $\frac{\epsilon}{2}$-پوشش از B_m و $\{v_1, \dots, v_K\}$ یک ϵ-پوشش از B_n. $\left\{ \begin{bmatrix} u_i \\ v_j \end{bmatrix} : \begin{matrix} 1\leq i \leq L \\ 1 \leq j \leq K \end{matrix} \right\}$ یک پوشش برای مجموعه‌ی

$B_m^{(1)} \times B_n^{(1)}$ است. فرض کنید $(u,v) \in B_m \times B_n$

$$\exists i,j : \begin{cases} \|u_i - u\| \le \frac{\epsilon}{2} \\ \|v_i - v\| \le \frac{\epsilon}{2} \end{cases} \Rightarrow d\left(\begin{bmatrix}u\\v\end{bmatrix}, \begin{bmatrix}u_i\\v_i\end{bmatrix}\right) \le \epsilon$$

$$\Rightarrow \mathcal{N}(B_m \times B_n, d, \epsilon) \le LK \le \left(\frac{6}{\epsilon}\right)^{m+n} \Rightarrow E\|A\| \lesssim \frac{\sigma}{1-\epsilon}\sqrt{(m+n)\log(1/\epsilon)} \quad ; \quad \epsilon = \frac{1}{2}$$

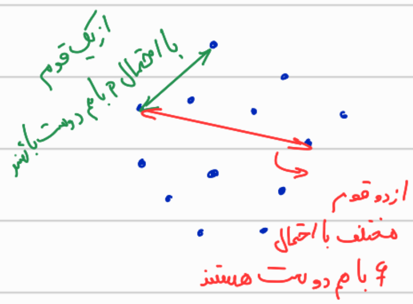
$$\lesssim \sigma\sqrt{m+n} \lesssim \sigma \max(\sqrt{m}, \sqrt{n})$$

جلسه چهاردهم: 1403, 8, 20

← خوشه‌بندی با استفاده از تجزیه طیفی

یادآوری: اگر A یک ماتریس تصادفی با درایه‌های مستقل و زیرگوسی σ با میانگین صفر باشد آنگاه $E\|A\| \le C.\sigma.\sqrt{m+n}$.

تعریف: گراف‌های تصادفی بلوکی (Stochastic Block Model)



گراف دوستی بین فارس‌ها و ترک‌ها

$$\text{ماتریس همسایگی } B = \begin{array}{c} \\ i \end{array}\begin{pmatrix} & & \overset{j}{\vdots} & \\ \cdots & \cdots & 1 & \\ & & & \\ & & & 0 \end{pmatrix} \qquad \begin{array}{l} (i \sim j) \to 1 \\ (i \nsim j) \to 0 \\ \text{قطر اصلی} = 0 \end{array}$$

فرض کنید جهت سادگی جمعیت فارس‌ها و ترک‌ها برابر باشند (هر کدام $\frac{n}{2}$). فرض کنید دسته‌ها مشخص باشد. در این صورت

$$EB = \begin{array}{c} \text{فارس} \\ \text{ترک} \end{array}\overset{\text{فارس} \quad \text{ترک}}{\begin{pmatrix} p & q \\ q & p \end{pmatrix}} \triangleq D \Rightarrow \underbrace{D + pI}_{\text{ماتریس رتبه } 2} = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

قطر اصلی = 0

$$(D+pI)\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix} = (p+q)\frac{n}{2}\begin{pmatrix}1\\1\\ \vdots\\1\end{pmatrix} \Rightarrow \lambda_1 = \frac{n(p+q)}{2}$$

$$(D+pI)\begin{pmatrix}1\\1\\ \vdots\\ -1\\ \vdots\\ -1\end{pmatrix} = \frac{p-q}{2}.n\begin{pmatrix}1\\1\\ \vdots\\1\\-1\\ \vdots\\-1\end{pmatrix} \Rightarrow \lambda_2 = \frac{p-q}{2}.n$$

← با هم ریختن سطرهای ماتریس فقط جای 1 و -1 ها در بردار ویژه عوض می‌شود ← میتوان خوشه‌بندی کرد

در واقعیت یک تحقق از B داریم و به میانگین دسترسی نداریم. فرض کنید q دو p باشد.

B: مقادیر ویژه $\longrightarrow$ بردارهای ویژه $\longrightarrow$ بردار ویژه دوم را در نظر بگیریم $\longrightarrow \begin{pmatrix} 1.7 \\ -0.3 \\ 2.34 \\ -17.08 \\ 0.07 \\ -0.011 \end{pmatrix} \xrightarrow{\text{sign}} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$

$$\vec{B - EB} = \begin{pmatrix} & & \\ & B_{ij} - EB_{ij} & \\ & & \end{pmatrix}_{n\times n}$$

$B_{ij} \in \{0,1\} \Rightarrow B_{ij} \sim subg(\frac{1}{4})$, $R_{ij} = B_{ij} - EB_{ij}$ میانگین صفر

مشابه مطالب جلسه قبل : $E\|B - EB\| \lesssim C.\sqrt{n}$, $\|EB\| = \frac{p+q}{2}.n$

$B = EB + R \Longrightarrow (n - \sqrt{n}).c \leq \|B\| \leq (n + \sqrt{n}).c$ به نظرم اگر B به امیدریاضی خود نزدیک است

قضیه Weyl: اگر S و T دو ماتریس مربعی متقارن باشند آنگاه اگر $\lambda_1(S) > \dots > \lambda_n(S)$ و $\lambda_1(T) > \dots > \lambda_n(T)$ داریم

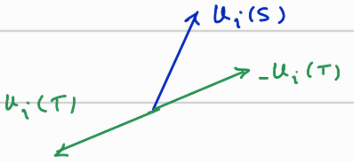
$$|\lambda_i(S) - \lambda_i(T)| \leq \|S - T\|_{op}$$

قضیه Davis-Kahan: فرض کنید S و T دو ماتریس دلخواه مربعی متقارن باشند و بردار ویژه‌ی iام متناظر با S و T مدنظر است. $(u_i(S), u_i(T))$

فرض کنید $\delta = \min\limits_{j \neq i} |\lambda_i(S) - \lambda_j(S)| > 0$ آنگاه (بردارها نرمالیزه باشند)

$$\sin \overbrace{\measuredangle\, u_i(S), u_i(T)}^{\text{زاویه بین بردارها}} \leq C.\frac{\|S - T\|}{\delta}$$

نتیجه $\Longrightarrow |u_i(S) - \theta u_i(T)| \leq \tilde{C}.\frac{\|S-T\|}{\delta}$; $\exists\, \theta \in \{1, -1\}$ (Matrix Analysis (Bhatia))



حال به مسئله‌ی اصلی برمی‌گردیم:

$S \longleftarrow B$; $u_2(S) = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ -1 \\ \vdots \\ -1 \end{pmatrix} \frac{1}{\sqrt{n}}$; $\delta = \min\limits_{j \neq 2} |\lambda_j(S) - \overset{\frac{p-q}{2}n}{\lambda_2(S)}|$

$T \longleftarrow EB$ $\quad u_2(T) = \dots$ $\quad = \underbrace{\min(q, \frac{p-q}{2})}_{\tau > 0}.n = \tau.n$

Davis-Kahan $\Longrightarrow |u_2(S) - \theta u_2(T)| \leq \frac{C.\sqrt{n}}{\tau n} = \frac{c}{\tau\sqrt{n}}$; $\exists\, \theta \in \{1, -1\}$

دقت کنید $E\|S-T\| \lesssim \sqrt{n}$ پس طبق مارکوف $\|S-T\| < 10\sqrt{n}$ با احتمال 0.90. یعنی رابطه بالا با احتمال بالا برقرار است.

$$|u_2(S) - u_2(T)| = \sqrt{\sum (u_{2k}(S) - u_{2k}(T))^2}$$

$$\geq \sqrt{\sum \frac{1}{n} \mathbb{1}\{u_{2k}(s) \cdot u_{2k}(T) < 0\}} = \frac{1}{\sqrt{n}} \sqrt{d_H(\sqrt{n}\, u_2(s),\ \mathrm{sign}(u_2(T)))}$$

$\Rightarrow \quad d_H(\sqrt{n}\ u_2(s),$ علامت بردار کشف شده $) \leq \frac{c^2}{\tau^2} = \frac{\tilde{c}}{\tau^2} \rightarrow$ یک تعداد ثابت درایه اختلاف وجود دارد

یک کاربرد دوم از قضیه فرآیندهای لیپ شیتز: ارتباط بین تخمین تجربی (توزیع تجربی) و توزیع واقعی بپردازیم.

$P_X : [0,1]$ یا $\mathbb{R}$ روی یک توزیع دلخواه

$$X_1, \dots, X_n \Rightarrow P_{emp}(x) = \frac{1}{n} \sum \mathbb{1}\{X_i = x\}$$

فاصله بین P_X و P_{emp} چطور است؟

$TV = \frac{1}{2} \sum |P_{emp}(x) - P_X(x)| = \sup_S |P_{emp}(S) - P(S)| \rightarrow$ می‌تواند 1 باشد.

Kolmogorov distance: $d_{Kol}(P, Q) = \sup_{t \in \mathbb{R}} |P(X \leq t) - Q(X \leq t)| \rightarrow$ P و Q روی $\mathbb{R}$ تعریف شده‌اند

$$D_{KL}(P \| Q) = \sum p(x) \log \frac{p(x)}{q(x)} \Rightarrow D_{KL}(P_X \| P_{emp}) = \infty!$$

IPM: $d(P, Q)$

$\mathcal{F} = \{f_1, f_2, \dots\} \Rightarrow$ P نزدیک Q: اگر $d_{\mathcal{F}}(P, Q) = \sup_{f \in \mathcal{F}} E_P[f] - E_Q[f]$ کم باشد.

مثال: Kolmogorov: $f_t(x) = \mathbb{1}\{x \leq t\}$

$W_1(P, Q) = \sup_{f:\ 1\text{-Lip}} E_P[f] - E_Q[f]$ در جلسه آینده واسرشتاین را بررسی می‌کنیم.

$$E\, W_1(P, P_{emp}) \xrightarrow{n \to \infty} 0 \quad \dots \quad \begin{cases} n^{-\frac{1}{3}} \rightarrow & \text{جلسه بعد ثابت می‌شود} \\ \dots & \\ n^{-\frac{1}{2}} \rightarrow \text{tight} \rightarrow & \text{در جلسات آینده مطرح می‌شود} \end{cases}$$

جلسه پانزدهم 22، 8، 1403

* کران روی فاصله‌ی واسرشتاین بین توزیع تجربی و واقعی
* روش زنجیره‌ای برای میانگین سوپریمم

$$W_1(P,Q) = \sup_{f:\,1\text{-Lip}} E_P(f) - E_Q(f)$$

$$\underset{X_1,\dots,X_n}{P_X} \;;\; P_X \,,\, P_{emp}^{(n)}(a) = P_n(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = a\}$$

$$E[W_1(P,P_n)] = E\Big[\sup_{f:\,1\text{-Lip}} E_P(f) - E_{P_n}(f)\Big]$$

$$= E\Big[\sup_{f:\,1\text{-Lip}} E_P(f(X)) - \frac{1}{n}\sum f(X_i)\Big]$$

$$= E\Big[\sup_{f} \overbrace{\frac{1}{n}\sum\big(E_P(f(X_i)) - f(X_i)\big)}^{Z_f}\Big]$$

فرض کنید از اینجا به بعد $\mathrm{supp}(X) = [0,1]$. آیا Z_f زیرگوسی و لیپ‌شیتز است؟

ادعا: $E[W_1(P,P_n)] = E\Big[\sup_{\substack{f:\,1\text{-Lip}\\ [0,1]\to[0,1]}} Z_f\Big]$ $\rightarrow$ چون f، 1-Lip است حداکثر تغییرات آن حداکثر یک واحد می‌شود

و چون Z_f با شیفت تغییری نمی‌کند می‌توان برد تابع را

نیز $[0,1]$ درنظر گرفت

زیرگوسی بودن Z_f :

$$Z_f = \frac{1}{n}\sum \underbrace{\big[E f(X_i) - f(X_i)\big]}_{R_f=[0,1] \text{ چون } \frac{1}{4}\text{-subg}} \Rightarrow Z_f \sim \frac{1}{4n}\ \text{subg}$$

لیپ‌شیتز بودن Z_f :

$$Z_f - Z_g \le C \cdot d(f,g) \qquad d(f,g) := \|f-g\|_\infty = \max_{x\in(0,1)} |f(x)-g(x)|$$

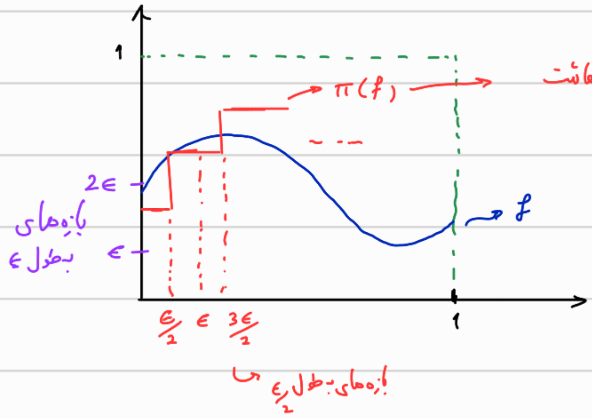
$$Z_f - Z_g = \big(E_P[f] - E_P[g]\big) - \frac{1}{n}\sum\big(f(X_i) - g(X_i)\big)$$

$$|Z_f - Z_g| \le |E_P(f-g)| + \frac{1}{n}\sum |f(X_i) - g(X_i)|$$

Jenson

$$\le E_P|f-g| + \frac{1}{n}\sum|f(X_i)-g(X_i)| \le 2\|f-g\|_\infty$$

محاسبه تعداد پوشش ها:



درمحور y به وسط بازه نگاشت می‌دهیم

$\Rightarrow \quad |\pi(f) - f| \le \epsilon$

تعداد بازه‌های روی محور x

تعداد نقاط شروع $= \frac{1}{\epsilon} \Rightarrow$ تعداد توابع پوشش دهنده $\le \frac{1}{\epsilon} \cdot 3^{\frac{2}{\epsilon}}$ ⟶ در هر بازه چون f، 1-لیپ‌شیتز است حداکثر 3 تابع جدید در بازه جدید ممکن است پدید بیاید

درمجموع: $\mathcal{N}(\mathcal{T}, \epsilon) \le e^{c/\epsilon}$

$$\Rightarrow \quad E[W_1(P, P_n)] \le 2\epsilon + \sqrt{2 \cdot \frac{1}{4n} \cdot \frac{c}{\epsilon}}$$

اگر $\epsilon = n^{-1/3}$: $E[W_1(P, P_n)] \le 2n^{-1/3} + \sqrt{c \cdot \frac{1}{n^{2/3}}} = O(n^{-1/3})$

* در ادامه درس خواهیم دید $E[W_1] \le O(n^{-1/2})$
* درفضای d بعدی روش فوق به کران زیر می‌رسد

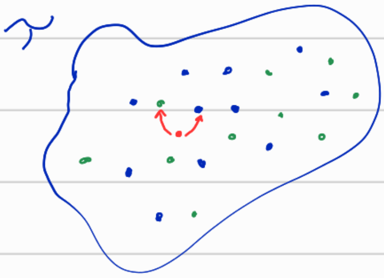
$$E[W_1] = O(n^{-\frac{1}{d+1}})$$

* مرور روش قبلی:

خطای ناشی از تقریب زدن

$$E[\sup X_t] \le E(c)\cdot\epsilon + E[\sup_{t \in \mathcal{N}_\epsilon} X_t]$$

* روش زنجیری:



نزدیک‌ترین نقطه بین آبی‌ها به t

نزدیک‌ترین نقطه بین سبزها به t

$$\sup X_t = \sup X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{\pi_2(t)} + X_{\pi_2(t)}$$

جلسه شانزدهم : 27 ,8 ,1403

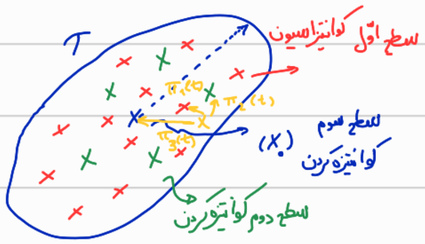
- انتگرال دادلی برای محاسبه امید سوپریمم
- مثال : کران بالا روی فاصله واسرشتاین

- فرآیند زیرگوسی: فضای متری $(\mathcal{T}, d)$. $\{X_t\}_{t\in\mathcal{T}}$. این فرآیند زیرگوسی است اگر $E\left[e^{\lambda(X_t - X_s)}\right] \leq e^{\frac{\lambda^2}{2} d(s,t)^2}$. $\begin{cases}\forall s,t\in\mathcal{T}\\ \lambda\in\mathbb{R}\end{cases}$

به عبارت دیگر $X_t - X_s$ یک متغیر تصادفی زیرگوسی با پارامتر $d(s,t)^2$ است.

نکته 1: $Pr\left[|X_t - X_s| > c\, d(s,t)\right] \leq 2e^{-\frac{c^2}{2}} \Rightarrow$ فرآیند با احتمال بالا لیپ شیتز است امّا در واقع ممکن است لیپ شیتز به معنای واقعی نباشد.

- روش Chaining :



$$\sup X_t = \sup X_t - X_{\pi_1(t)} + X_{\pi_1(t)} - X_{\pi_2(t)} + X_{\pi_2(t)} - X_{\pi_3(t)} + X_{\pi_3(t)}$$

$$\leq \sum_{k=0}^{2} \sup\left(X_{\pi_k(t)} - X_{\pi_{k+1}(t)}\right) + X_{\pi_3(t)} \rightarrow t = \pi_0(t)$$ با تعریف

$$\Rightarrow E \sup X_t \leq E \sum_{k=0}^{2} \sup\left(X_{\pi_k(t)} - X_{\pi_{k+1}(t)}\right) + E X_0$$

فرض کنید $\mathcal{T}$ مجموعه متناهی باشد. $\mathcal{N}_k$: یک پوشش برای مجموعه t در شعاع $\epsilon = 2^{-k}$ می‌گیریم $(k\in\mathbb{Z})$

فرض کنید k_{min} و k_{max} وجود دارد. $\mathcal{N}_{k_{min}} : \exists x_{t_0} : d(x_t, x_{t_0}) < 2^{-k_{min}} \quad \forall t\in\mathcal{T}$

$$\mathcal{N}_{k_{max}} : \pi_{k_{max}}(t) = t \;;\; \forall t\in\mathcal{T}$$

پس می‌توان گفت

$$X_t = \sum_{j=k_{min}+1}^{k_{max}} X_{\pi_j(t)} - X_{\pi_{j-1}(t)} + \underbrace{X_{k_{min}}}_{X_{t_0}}$$

حال به محاسبه‌ی $E[\sup ...]$ برمی‌گردیم. $E\left[\sup X_t\right] = E\left[\sup \sum ...\right] + E X_{t_0}$

فرض کنید فرآیند دارای میانگین صفر باشد. در این صورت $E X_{t_0} = 0$.

$$\Rightarrow E\left[\sup_{t\in\mathcal{T}} X_t\right] \leq \sum E\left[\sup\left(X_{\pi_j(t)} - X_{\pi_{j-1}(t)}\right)\right]$$

طبق فرض $X_{\pi_j(t)} - X_{\pi_{j-1}(t)}$ ، $d(\pi_j(t),\pi_{j-1}(t))^2$ زیرگوسی است.

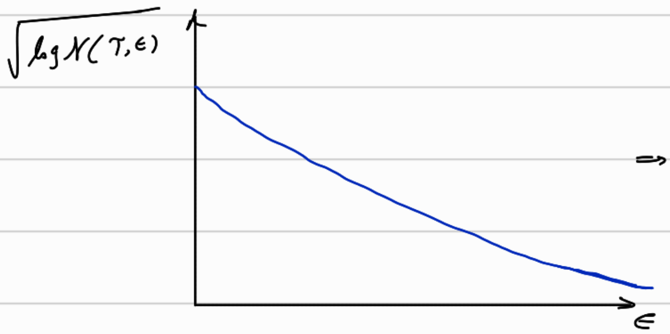
$$d(\pi_j(t),\pi_{j-1}(t)) \leq d(\pi_j(t),t) + d(\pi_{j-1}(t),t) \leq 2^{-j} + 2^{-(j-1)} = 3.2^{-j}$$

$$E\left[\sup_{\substack{\pi_j(t)\in N_j\\ \pi_{j-1}(t)\in N_{j-1}}} X_{\pi_j(t)} - X_{\pi_{j-1}(t)}\right] \leq \sqrt{2.(3.2^{-j})^2 . \log(\underbrace{|N_j|.|N_{j-1}|}_{\leq |N_j|^2})}$$

چون $|N_j|\geq|N_{j-1}|$ $\rightarrow$

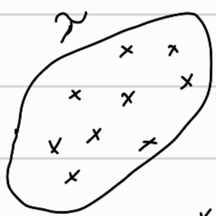
$$\leq c.\,2^{-j}\sqrt{\log|N_j|}$$

$$\Rightarrow \sum_{k_{min}+1}^{k_{max}} E[..] \leq c\sum_{j=k_{min}+1}^{k_{max}} 2^{-j}\sqrt{\log|N_j|} \leq c.\sum_{j=-\infty}^{\infty} 2^{-j}\sqrt{\log|N_j|}$$ ؛ باند دادلی



$$\Rightarrow c\sum_{j=-\infty}^{\infty} 2^{-j}\sqrt{\log|N_j|} \leq 2c.\underbrace{\int_0^{\infty}\sqrt{\log|N(\mathcal{T},\epsilon)|}\,d\epsilon}_{\text{انتگرال دادلی}}$$

فرض کنید $\mathcal{T}$ نامحدود باشد



چگال شمارا

$$\mathcal{T}_0 \subseteq \mathcal{T};\quad E\left[\sup_{\mathcal{T}} \dots\right] = E\left[\sup_{\mathcal{T}_0}\dots\right]$$

خاصیت همگرایی یکنوا

$$X_1, X_2, \dots :\ E\sup X_i = E\left[\lim_{N\to\infty}\max_{1\leq i\leq N} X_i\right] = \lim_{N\to\infty} E\left[\max_{1\leq i\leq N} X_i\right]$$

مثال: $X_1,\dots,X_n$; $P^{(n)}_{emp}$

$$E[W_1(P,P_n)] = E\left[\sup_{f:\,1\text{-Lip}} \underbrace{E_P[f] - E_{P_n}[f]}_{Z_f}\right] \rightarrow d(f,g) = \|f-g\|_\infty$$

فضای توابع 1-Lip را با متر $\|f-g\|_\infty$ درنظر گرفتیم.

$$Z_f = E_P[f] - \frac{1}{n}\sum f(X_i) \Rightarrow |Z_f - Z_g| \leq 2\|f-g\|_\infty$$

$$Z_f - Z_g \leq \frac{\|f-g\|_\infty}{\sqrt{n}}$$ به‌طور حسی دیدیم با توجه به واریانس با احتمال بالا

$$Z_f - Z_g = \frac{1}{n} \sum \left(g(X_i) - f(X_i) - \underset{X\sim P}{E}[g(X) - f(X)] \right)$$

$|g(X_i) - f(X_i)| \leq \|g-f\|_\infty \Rightarrow$ $g(X_i)-f(X_i)$ زیرگوسی با $\|g-f\|_\infty^2$ است

$$\Rightarrow Z_f - Z_g \sim \frac{\|f-g\|_\infty^2}{n}\text{-subg} \Rightarrow d(f,g) = \frac{\|f-g\|_\infty}{\sqrt{n}}$$

$$\Rightarrow E\, W_1(P, P_n) \leq c \int_0^\infty \sqrt{\log N\left(1\text{-Lip}\text{فضای}, \epsilon, \frac{\|f-g\|_\infty}{\sqrt{n}}\right)}\, d\epsilon$$

یادآوری: $\log N\left(1\text{-lip فضای}, \| \ \|_\infty, \epsilon\right) = \begin{cases} \frac{c}{\epsilon} & ; \epsilon < \frac{1}{2} \\ 0 & ; \epsilon \geq \frac{1}{2} \end{cases}$

$f: [0,1] \to [0,1]$

$$\Rightarrow E\, W_1(P, P_n) \lesssim \int_0^\infty \sqrt{\log N\left(\dots, \frac{\| \ \|_\infty}{\sqrt{n}}, \epsilon\right)}\, d\epsilon = \int_0^\infty \sqrt{\log N(\dots, \| \ \|_\infty, \underbrace{\epsilon\sqrt{n}}_{\tilde{\epsilon}})}\, d\epsilon$$

$$= \frac{1}{\sqrt{n}} \int_0^{\frac{1}{2}} \underbrace{\sqrt{\log N(\dots, \| \ \|_\infty, \tilde{\epsilon})}}_{\frac{c}{\sqrt{\epsilon}}}\, d\tilde{\epsilon} = \frac{c'}{\sqrt{n}}$$

می‌توان نشان داد برای $d \geq 3$، باند بالا $n^{-1/d}$ و برای $d=2$، $\frac{\log n}{\sqrt{n}}$ می‌شود.

جلسه هفدهم: $1403,8,29$

- کران پایین برای امید سوپریمم
- نرآینه‌های گوسی
- قضیه Sudakov

تمرین: اگر $X_1, \dots, X_n$، n متغیر تصادفی گوسی میانگین صفر و واریانس σ^2 (iid) باشند آنگاه

$$c' \sqrt{\sigma^2 \log n} \leq E[\max X_i] \leq \sqrt{2\sigma^2 \log n}$$

پس برای متغیرهای گوسی $E[\max X_i] = \Theta(\sigma\sqrt{\log n})$.

فرآیندهای گوسی: $\{X_t\}_{t\in T}$ اگر برای هر $t_1,\dots,t_n$، n تایی $(X_{t_1},\dots,X_{t_n})$ مشترکاً گوسی باشند. (بازای هر $n\in \mathcal{N}$ برقرار است)

نکته: فرض کنید فرآیند دارای میانگین صفر باشد.

$$E\left[e^{\overset{\sim N(0,Var(X_t-X_s))}{\lambda(X_t-X_s)}}\right] \le \exp\left(\frac{\lambda^2}{2}Var(X_t-X_s)\right)$$

$$= \exp\left(\frac{\lambda^2}{2}E[(X_t-X_s)^2]\right) = \exp\left(\frac{\lambda^2}{2}\overbrace{\|X_t-X_s\|_2^2}^{d(s,t)^2}\right)$$

فرآیند گوسی، زیرگوسی نیز هست. $\Rightarrow$ فاصله‌ی مشخصه را اینگونه تعریف کنیم $\rightarrow$ $d_{ch}^2(s,t)=\|X_t-X_s\|_2^2$

مثال: فرض کنید $Z\sim \mathcal{N}(0,I_d)$ باشد. یک فرآیند تصادفی با اندیس t $\rightarrow$ $X_t=\langle t,Z\rangle$

می‌خواهیم EX_tX_s و $\|X_t-X_s\|_2^2$ را محاسبه کنیم.

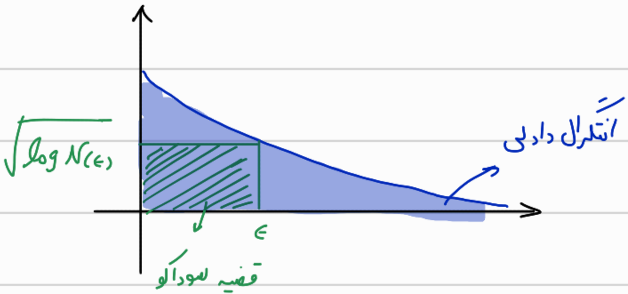
$$EX_tX_s = E\{t^TZ\,.\,Z^Ts\} = t^T\overbrace{E(ZZ^T)}^{I_d}s = t^Ts = \langle t,s\rangle$$

$$\|X_t-X_s\|_2^2 = E((X_t-X_s)^2) = d_{ch}^2(s,t) \Longrightarrow d_{ch}^2(s,t)=\|t-s\|_2^2$$

قضیه Sudakov (کران پایین برای امید ریاضی سوپریمم)

یادآوری: فرض کنید X_tها تشکیل یک فرآیند زیرگوسی دهند.

$$E\left[\sup_{t\in T} X_t\right] \lesssim \int_0^\infty \sqrt{\log \mathcal{N}(T,d,\epsilon)}\,d\epsilon$$

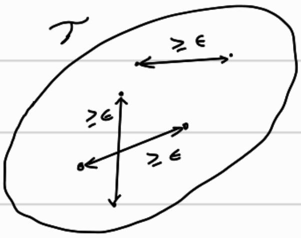


Sudakov: برای فرآیند گوسی میانگین صفر با فاصله مشخصه d_{ch}،

$$E\left[\sup_{t\in T} X_t\right] \ge \sup_{\epsilon>0} c\cdot\epsilon\sqrt{\log \mathcal{P}(T,d_{ch},\epsilon)} \ge \sup_{\epsilon>0} c\cdot\epsilon\sqrt{\log \mathcal{N}(T,d_{ch},\epsilon)}$$

اثبات Sudakov:

مجموعه ϵ-packing متقابل را با Q نمایش دهید.



$$E[\sup_{t\in T} X_t] \geq E[\max_{t\in Q} X_t]$$

اگر فرض کنید $\{X_t\}_{t\in Q}$ از هم مستقل باشند. در این صورت $E[\max_{t\in Q} X_t] \geq c\epsilon\sqrt{\log P(\epsilon)}$. $d^2(s,t) = E((X_t-X_s)^2) \geq \epsilon^2$

واریانس X_t در حدود $X_t - X_{t_0}$ است $\Longrightarrow$ $t_0\in Q:\ E[\max_{t\in Q} X_t] = E[\max_{t\in Q} X_t - X_{t_0}] + \overbrace{E[X_{t_0}]}^{0\ :\ \text{میانگین صفر بودن}}$

چون X_t ها مستقل نیستند روش بالا کاملاً نادرست است.

لم (مقایسه‌ای Slepian): فرض کنید n متغیر تصادفی گوسی $X_1, \dots, X_n$ و n متغیر تصادفی گوسی $Y_1, \dots, Y_n$ داده شده باشند (میانگین صفر)

$$\forall i,j:\ E[(X_i-X_j)^2] \geq E[(Y_i-Y_j)^2] \quad \text{فرض کنید} \Longrightarrow \quad E[\max X_i] \geq E[\max Y_i]$$

در جلسه آینده این لم ثابت می‌شود. فعلاً با استفاده از این لم به ادامه اثبات اصلی می‌پردازیم. برای مجموعه Q:

$$\forall (s,t):\ d_{ch}^2(s,t) > \epsilon^2 \longrightarrow E[(X_t - X_s)^2] > \epsilon^2$$

فرض کنید $Y_t\ (t\in Q)$ را متغیرهای iid به شکل $\mathcal{N}(0, \frac{\epsilon^2}{2})$ بگیریم و Y_t ها مستقل باشند.

$E[(Y_t-Y_s)^2] = Var(Y_t) + Var(Y_s) = \epsilon^2 \rightarrow$ شرایط Slepian برقرار است.

$$\Longrightarrow E[\max_{t\in Q} X_t] \geq E[\max_{t\in Q} Y_t] \geq c\cdot\sqrt{\frac{\epsilon^2}{2}\log P(\epsilon)} \geq c'\epsilon\sqrt{\log P(T,d,\epsilon)}$$

کاربرد: دیدیم اگر $A = \begin{pmatrix} & \\ & \end{pmatrix}_{m\times n}$ ، درایه‌ها زیرگوسی و مستقل باشند $E\|A\|_{op} \lesssim c\cdot\sigma\sqrt{m+n}$

فرض کنید $A_{ij}\sim\mathcal{N}(0,\sigma^2)$ و مستقل باشند. آیا $E\|A\|_{op} \gtrsim c\sigma\sqrt{m+n}$ ؟

$$E\|A\|_{op} = E[\sup_{u,v} \overbrace{u^T A v}^{Y_{u,v}}]$$

آیا $\{Y_{u,v}\}$ فرآیند گوسی است؟

$Y_{u,v} = \sum_{i,j} u_i v_j A_{ij} \rightsquigarrow$ فرآیند گوسی است

$$\Rightarrow \quad E\left[\sup Y_{u,v}\right] \geq c\epsilon \sqrt{\log N(B_m \times B_n, d_{ch}, \epsilon)}$$

یادآوری: در کران بالا $d\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix}\right) = \|u-\tilde{u}\| + \|v-\tilde{v}\|$. اینجا باید d_{ch} را محاسبه کنیم.

$$d_{ch}\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix}\right) = E\left[(Y_{uv} - Y_{\tilde{u}\tilde{v}})^2\right] = E\left[(u^T A v - \tilde{u}^T A \tilde{v})^2\right]$$

$$= E\left[\left(\sum_{i,j} (u_i v_j - \tilde{u}_i \tilde{v}_j) A_{ij}\right)^2\right]$$

چون $EA_{ij}=0$

$$= \sum_{i,j} \operatorname{Var}\left((u_i v_j - \tilde{u}_i \tilde{v}_j) A_{ij}\right)$$

$$= \sigma^2 \sum_{i,j} (u_i v_j - \tilde{u}_i \tilde{v}_j)^2$$

$$= \sigma^2 \sum_{i,j} (u_i^2 v_j^2 + \tilde{u}_i^2 \tilde{v}_j^2 - 2 u_i \tilde{u}_i v_j \tilde{v}_j)$$

$$= \sigma^2 \left(\underbrace{\|u\|^2 \|v\|^2}_{1} + \underbrace{\|\tilde{u}\|^2 \|\tilde{v}\|^2}_{1} - 2\langle u, \tilde{u}\rangle \langle v, \tilde{v}\rangle\right)$$

$$= 2\sigma^2 \left(1 - \langle v, \tilde{v}\rangle \langle u, \tilde{u}\rangle\right)$$

اگر فرض کنیم $\sigma = 1$ آنگاه

$$d_{ch}^2\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix}\right) = 2\left(1 - \langle u, \tilde{u}\rangle \langle v, \tilde{v}\rangle\right) \geq \underbrace{2(1 - \langle u, \tilde{u}\rangle)}_{\|u-\tilde{u}\|_2^2}, \underbrace{2(1 - \langle v, \tilde{v}\rangle)}_{\|v-\tilde{v}\|_2^2}$$

$$\geq \max\left(\|u-\tilde{u}\|_2^2, \|v-\tilde{v}\|_2^2\right) = d_{max}^2$$

$$N(B_m \times B_n, d_{ch}, \epsilon) \geq N(B_m \times B_n, d_{max}, \epsilon)$$

$$Q_1, \dots, Q_L : \forall p \ \exists \ell \ \ d_{ch}(P, Q_\ell) < \epsilon \Rightarrow d_{max} < \epsilon \checkmark$$

فرض کنید $\begin{bmatrix} u_1 \\ v_1 \end{bmatrix}, \dots, \begin{bmatrix} u_\ell \\ v_\ell \end{bmatrix}$ پوشش $B_m \times B_n$ باشد. در این صورت

$$\forall (u,v)\ ;\ \exists \ell \quad d_{max}\left(\begin{bmatrix}u\\ v\end{bmatrix},\begin{bmatrix}u_\ell\\ v_\ell\end{bmatrix}\right)<\epsilon \Rightarrow \|u-u_\ell\|<\epsilon,\ \|v-v_\ell\|<\epsilon \Rightarrow \begin{cases}\{u_1,\dots,u_\ell\}: B_m \text{ پوشش}\\ \{v_1,\dots,v_\ell\}: B_n \text{ پوشش}\end{cases}$$

$$\Rightarrow \log L > \max(m,n)\log\left(\tfrac{1}{\epsilon}\right)$$

$$\Rightarrow E\|A\| \ge c.\,\epsilon\sqrt{\max(m,n)\log\left(\tfrac{1}{\epsilon}\right)} \overset{\epsilon=\frac{1}{e}}{=} c'\sqrt{\max(m,n)} \ge c''\sqrt{m+n}$$

جلسه هجدهم: 4، 8، 1403

- اثبات لم Slepian

- لم مقایسه‌ای: اگر $(X_1,\dots,X_n)$ یک دسته متغیرهای تصادفی مشترکاً گوسی (میانگین صفر) و $(Y_1,\dots,Y_n)$ نیز همین طور باشند و (X و Y مستقل)

$$\forall i,j: E\left[(X_i-X_j)^2\right] \ge E\left[(Y_i-Y_j)^2\right] \Rightarrow E\left[\max X_i\right] \ge E\left[\max Y_i\right]$$

- ایده 1: فرض کنید $f:\mathbb{R}^n \to \mathbb{R}$ یک تابع دلخواه باشد و $h(t)=E\left[f(\sqrt{t}\,X+\sqrt{1-t}\,Y)\right]\ ;\ t\in(0,1)$

اگر h تابع صعودی باشد: $Ef(X) \ge Ef(Y) \Leftarrow h(1)>h(0) \Leftarrow \forall t: h'(t)>0$

پس اگر $f(x)=\max x_i$ ها آنگاه به نتیجه مورد نظر می‌رسیم. اما نمی‌خواهیم max را قرار دهیم زیرا مشتق‌پذیر نیست.

- ایده 2: برای اینکه از max استفاده نکنیم می‌توانیم از جایگزین آن یعنی $f_\beta(\vec{x})=\frac{1}{\beta}\log\sum e^{\beta x_i}$ استفاده کنیم:

$$\max x_i \le f_\beta(\vec{x}) \le \frac{1}{\beta}\log n.e^{\beta \max x_i} = \max x_i + \frac{\log n}{\beta}$$

پس با توجه به قضیه ساندویچ $\lim_{\beta\to\infty} f_\beta(\vec{x})=\max x_i$

$$h_\beta(t)=E\left[f_\beta(\sqrt{t}\,X+\sqrt{1-t}\,Y)\right]$$

نشان خواهیم داد که h_β تابعی صعودی از t است و درنتیجه $\forall \beta>0: E\left[f_\beta(X)\right] \ge E\left[f_\beta(Y)\right]$

کافیست نشان دهیم $h'_\beta(t)$ برای هر $t\in(0,1)$ نامنفی است.

* اتحاد درمورد توزیع های گوسی :

$$X \sim \mathcal{N}(0,1): \quad E[Xf(X)] = E[f'(X)]$$

$$X \sim \mathcal{N}(0,I),\ f:\mathbb{R}^n \to \mathbb{R} \Rightarrow E[Xf(X)] = \begin{pmatrix} E(X_1 f(X)) \\ \vdots \\ E(X_n f(X)) \end{pmatrix} = \begin{pmatrix} E[\frac{\partial f}{\partial x_1}] \\ \vdots \\ E[\frac{\partial f}{\partial x_n}] \end{pmatrix} = E[\nabla f(X)]$$

$$X \sim \mathcal{N}(0,C): \quad E[Xf(X)] = C\,E[\nabla f(X)]$$

اثبات: $X = C^{1/2} Z = C^{1/2} E[Z \overbrace{f(C^{1/2} Z)}^{g(Z)}] = C^{1/2} E[\nabla g(Z)] = C\,E[\nabla f(X)]$

ستون i ام C چون C متقارن است

$$E[X_i f(X)] = \langle C_i, E[\nabla f(X)] \rangle$$

حال به مشتق $h_\beta(t)$ برمی گردیم.

$$h_\beta(t) = E[f_\beta(\sqrt{t}\,X + \sqrt{\bar{t}}\,Y)]$$

$$h'_\beta(t) = E\left[\frac{d}{dt} f_\beta(\overbrace{\sqrt{t}\,X + \sqrt{\bar{t}}\,Y}^{Z})\right] = E\left\langle \frac{1}{2\sqrt{t}}X - \frac{1}{2\sqrt{\bar{t}}}Y,\ \nabla f_\beta(\sqrt{t}\,X + \sqrt{\bar{t}}\,Y) \right\rangle$$

به منظور مولفه i ام گرادیان است

$$= \frac{1}{2}\sum_{i=1}^{n} E\left[\left(\frac{X_i}{\sqrt{t}} - \frac{Y_i}{\sqrt{1-t}}\right) \underbrace{\frac{\partial f_\beta}{\partial z_i}(\sqrt{t}\,X + \sqrt{1-t}\,Y)}_{g_i(X,Y)}\right]$$

$$= \frac{1}{2}\sum_{i=1}^{n} E\left[\frac{1}{\sqrt{t}} \langle \nabla_X g_i(X,Y), C_i^X \rangle - \frac{1}{\sqrt{1-t}} \langle \nabla_Y g_i(X,Y), C_i^Y \rangle\right]$$

$$= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\left[\frac{1}{\sqrt{t}} C_{ij}^X \frac{\partial g_i}{\partial x_j} - \frac{1}{\sqrt{1-t}} C_{ij}^Y \frac{\partial g_i}{\partial Y_j}\right]$$

$$= \frac{1}{2}\sum_{i,j=1}^{n} E\left[\frac{\partial^2 f}{\partial z_i \partial z_j}(C_{ij}^X - C_{ij}^Y)\right] = \frac{1}{2} E\, Tr(\nabla^2 f_\beta (C_X - C_Y))$$

$$f_\beta = \frac{1}{\beta}\log \sum e^{\beta z_i} \Rightarrow \frac{\partial f_\beta}{\partial z_i} = \overbrace{\frac{e^{\beta z_i}}{\sum e^{\beta z_j}}}^{q_i} \ (\text{soft max})$$

$$\Rightarrow \frac{\partial^2 f_\beta}{\partial z_i \partial z_j} = \begin{cases} \dfrac{-\beta e^{\beta z_i} e^{\beta z_j}}{(\sum e^{\beta z_j})^2} = -q_i q_j \beta & ; i \neq j \\ -\beta q_i^2 + \beta q_i & ; i \neq j \end{cases}$$

$$h'_\beta(t) = \frac{1}{2} \sum_{i,j} E_{X,Y} \left[\underbrace{\frac{\partial^2 f_\beta}{\partial z_i \partial z_j} \left(\sqrt{t}\, X + \sqrt{\bar{t}}\, Y\right) \left(\tilde{x}_i \tilde{x}_j - \tilde{y}_i \tilde{y}_j \right)} \right]$$

$\tilde{x}, \tilde{y}$ ← متغیر هم‌توزیع و مستقل از X

Y 〃 〃 〃

$$\sum \frac{\partial^2 f_\beta}{\partial z_i \partial z_i} (\dots)\, \tilde{x}_i \tilde{x}_j = \beta \left[\sum_i q_i \tilde{x}_i^2 - \sum_i q_i^2 \tilde{x}_i^2 - \sum_{i,j:\, j \neq i} q_i q_j \tilde{x}_i \tilde{x}_j \right]$$

$$= \beta \Big[\underbrace{\sum_i q_i \tilde{x}_i^2 - \Big(\sum_i q_i \tilde{x}_i \Big)^2}_{Var(V)} \Big] \quad ; \quad V = \tilde{x}_i \text{ with probability } q_i$$

$$= \beta \left(\sum_{i,j} \frac{q_i q_j \tilde{x}_i^2 + q_i q_j \tilde{x}_j^2}{2} - q_i q_j \tilde{x}_i \tilde{x}_j \right)$$

$$= \frac{\beta}{2} \sum_{i,j} q_i q_j (\tilde{x}_i - \tilde{x}_j)^2$$

$$= \frac{\beta}{4} E_{X,Y} \sum_{i,j} \underbrace{q_i q_j}_{\geq 0} E_{\tilde{x},\tilde{y}} \underbrace{\left[(\tilde{x}_i - \tilde{x}_j)^2 - (\tilde{y}_i - \tilde{y}_j)^2 \right]}_{\geq 0 \,:\, \text{طبق فرض}} \geq 0$$

جلسه نوزدهم : 1403 ,8 ,6

- فرآیندهای تجربی
- فاصله کولموگروف و متقارن سازی

- واسرشتاین :

$$W_1(P, Q) = \sup_{f:\, 1\text{-Lip}} E_P(f) - E_Q(f)$$

$$X_1, \dots, X_n \sim P \;;\; W_1(P, P_n) = O\left(\frac{1}{\sqrt{n}}\right)$$

$$W_1(P, P_n) = \sup_{f:\, 1\text{-Lip}} E_P[f(X)] - \frac{1}{n} \sum f(X_i)$$

قضیه Gilvenko - Cantelli : $X_1, \dots, X_n$ به همراه توزیع تجربی P_n :

$$\sup_t \left| F_P(t) - F_{P_n}(t) \right| \xrightarrow[n \to \infty]{} 0$$

$F_P(t) = P(X \leq t)$ ، $F_{P_n}(t) = P_n(X \leq t) = \dfrac{\# \{ i : X_i \leq t \}}{n}$

Kolmogorov Distance: $Kol(P, Q) = \sup_{t \in \mathbb{R}} P(X \leq t) - Q(X \leq t)$

$$\mathcal{F}_{kol} = \left\{ 1\{ x \leq t \} \,;\, t \in \mathbb{R} \right\} : \quad kol(P, Q) = \sup_{f \in \mathcal{F}_{kol}} E_P(f) - E_Q(f)$$

$$GK\text{-}Thm: E\, kol(P, P_n) \xrightarrow[n \to \infty]{} 0$$

$$E\, kol(P, P_n) = E\Big[\sup_{f \in \mathcal{F}_{kol}} \underbrace{E_P[f] - \frac{1}{n} \sum f(X_i)}_{X_f} \Big]$$

برای باندزدن از قضیه دادلی استفاده می کنیم.

$d(f, g)$

$$X_f - X_g \sim \left(\frac{\|f - g\|_\infty}{\sqrt{n}} \right)^2 \rightarrow$$ مانند حالت واسرشتاین $\Rightarrow$ فرآیند زیرگوسی است

$$E\left[kol(P, P_n) \right] \lesssim \int_0^\infty \sqrt{\log N\left(F_{kol}, \frac{\|\ \|_\infty}{\sqrt{n}}, \epsilon\right)}\, d\epsilon$$

$1\{x \leq s\}$

$f_s, f_t \in \mathcal{F}_{kol} : \|f_s - f_t\|_\infty = 1\{s \neq t\} \Rightarrow$ اگر $\epsilon > 1$ ، $N(\epsilon) = 1$ است

$$= \int_0^\infty \sqrt{\log N(F_{kol}, \|\ \|_\infty, \epsilon\sqrt{n})}\, d\epsilon = \frac{1}{\sqrt{n}} \int_0^\infty \sqrt{\log N(F_{kol}, \|\ \|_\infty, \epsilon)}\, d\epsilon$$

اگر $\epsilon < 1$ باشد $\infty = N(\epsilon)$ می شود پس انتگرال دادلی بی نهایت می شود و به کار نمی آید.

به متقارن سازی :

$X_1, \dots, X_n \sim P$

$Y_1, \dots, Y_n \sim P$

$$\underset{X_1,\dots,X_n \sim P}{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n f(X_i) - E_P(f) \right] = E\left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \left(f(X_i) - E_P[f(Y_i)] \right) \right]$$

تعویض E, $\sup$

$$\leq E_{X^n, Y^n}\Big[\sup_f \frac{1}{n} \sum \overbrace{\left(f(X_i) - f(Y_i) \right)}^{Z_i} \Big]$$

$\epsilon^n : \epsilon_i$ iid

$\epsilon_i : \pm 1 ; p = \frac{1}{2}$

$$= E_{X^n, Y^n, \epsilon^n}\Big[\underbrace{\sup_f \frac{1}{n} \sum \epsilon_i \left(f(X_i) - f(Y_i) \right)}_{} \Big]$$

$$\leq \sup_f \frac{1}{n} \sum \epsilon_i f(X_i) + \sup_f \frac{1}{n} \sum (-\epsilon_i) f(Y_i)$$

$$\boxed{\leq 2\, E_{X^n, \epsilon^n}\left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \epsilon_i f(X_i) \right]}$$

تمرین: نشان دهید کران بالا از: $4E\left[\sup \frac{1}{n} \sum f(X_i) - E_P(f) \right]$.

هدف: $$E\left[\sup_{f \in \mathcal{F}_{kol}} \frac{1}{n} \sum f(X_i) - E_P(f) \right] \lesssim \frac{1}{\sqrt{n}} E_{X^n, \epsilon^n}\Big[\sup_{f \in \mathcal{F}_{kol}} \overbrace{\frac{1}{\sqrt{n}} \sum \epsilon_i f(X_i)}^{H_f} \Big]$$ حال مسئله اصلی بررسی کردیم.

اگر x_i ها تصادفی نباشند پس می‌توان گفت $\frac{1}{\sqrt{n}} \sum \epsilon_i f(x_i) \overset{subg}{\sim} \frac{1}{n} \sum_{i=1}^{n} f^2(x_i)$ پس

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i f(x_i) \overset{subg}{\sim} \|f(x)\|^2_{L^2(P_n)}$$

; یادآوری: $\|z^2\|^2_{L^2(\mu)} = E_{z\sim\mu}[z^2]$

آیا $H_f = \frac{1}{\sqrt{n}} \sum \epsilon_i f(x_i)$ زیرگوسی است؟

$$H_f - H_g = \frac{1}{\sqrt{n}} \sum \epsilon_i (f(x_i) - g(x_i)) \overset{subg}{\sim} \frac{1}{n} \sum (f(x_i) - g(x_i))^2 = \|f-g\|^2_{L^2(P_n)}$$

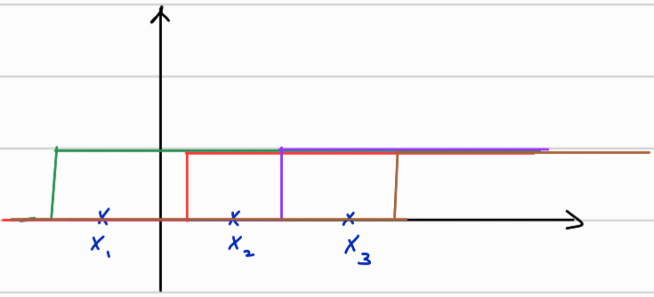
پس فرآیند با متر $d_n^2(f,g) = \|f-g\|^2_{L^2(P_n)}$ زیرگوسی می‌شود. پس طبق دادلی:

$$E_{\epsilon^n}\Big[\sup_{f\in F} H_f\Big] \overset{\text{دادلی}}{\lesssim} \int_0^1 \sqrt{\log N(F_{kol}, \|\ \|_{L^2(P_n)}, \epsilon)}\, d\epsilon$$

$$N(F_{kol}, \|\ \|_{L^2(P_n)}, \epsilon) \le N(F_{kol}, \|\ \|_{L^\infty(P_n)}, \epsilon) \quad *$$

$$\|f-g\|_{L^\infty(P_n)} = \max_{1\le i\le n} |f(x_i) - g(x_i)| \Longrightarrow \|f-g\|_{L^2(P)} \le \|f-g\|_{L^\infty(P)}$$

فقط در نقاط دیتاست بررسی می‌کنیم
و بقیه نقاط مهم نیستند



$$\|f_s - f_t\|_{L^\infty(P_n)} = \begin{cases} 1 & ; \ \exists X_i \in (s,t) \\ 0 & ; \ \text{O.W.} \end{cases}$$

→ اگر $n = 3$ ، این 4 تادونه یک
پوشش از توابع با $\|\ \|_{L^\infty(P_n)}$ است.

پس برای هر n ، $N(F, \|\ \|_{L^\infty(P_n)}, \epsilon) = n+1$ برای $0 < \epsilon < 1$. پس در نتیجه

$$E_{\epsilon^n}\Big[\sup_{f\in F_{kol}} H_f\Big] \lesssim \int_0^1 \sqrt{\log(n+1)}\, d\epsilon = \sqrt{\log(n+1)}$$

حال به هدف اولیه باز می‌گردیم:

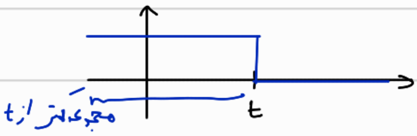
$$E\Big[\sup_{f\in F_{kol}} \frac{1}{n} \sum f(x_i) - E_P[f]\Big] \lesssim \frac{1}{\sqrt{n}} E_{x^n} \underbrace{E_{\epsilon^n}\Big[\sup_{f\in F_{kol}} H_f\Big]}_{\le \sqrt{\log(n+1)}} \lesssim \sqrt{\frac{\log n}{n}}$$

در بعد بالاتر:

$$d_{\mathcal{F}}(P,Q)=\sup_{f\in\mathcal{F}} E_P(f)-E_Q(f)$$

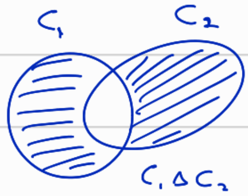
برای پیوسته بودن مبحث $\mathcal{C} \longleftrightarrow \mathcal{F}$. $\mathcal{C}$ یک خانواده از زیرمجموعه‌های $\mathbb{R}^d$ است. $\mathcal{F}_{\mathcal{C}}=\{1[x\in C]\,;\,C\in\mathcal{C}\}$.

برای مثال در کلموگروف $\mathcal{C}$ مجموعه بازه‌های یک طرفه بود.



$$E\left[d_{\mathcal{C}}(P,P_n)\right]=E\left[\sup_{C\in\mathcal{C}} P(X\in C)-\frac{1}{n}\sum 1[X_i\in C]\right]$$

$$\|f_{C_1}-f_{C_2}\|_{L^\infty(P_n)}=\begin{cases}1 & \exists X_i\in C_1\Delta C_2\\ 0 & \text{o.w.}\end{cases}$$



تفاضل متقارن: $C_1\Delta C_2$

$$\mathcal{N}(\mathcal{F},\|\ \|_\infty,\epsilon)\le\left|\mathcal{C}\cap\{X_1,\dots,X_n\}\,;\,C\in\mathcal{C}\right|$$

$\mathcal{C}\cap\{X_1,\dots,X_n\}=\{A_1,\dots,A_m\}$

$A_1=C_1\cap\{X_1,\dots,X_n\}$

$\vdots$

$A_m=C_m\cap\{X_1,\dots,X_n\}$

$A_i\neq A_j:\ \exists X_k:\ X_k\in C_i\Delta C_j \Rightarrow \|f_{C_i}-f_{C_j}\|_\infty=1 \Rightarrow$ $f_{C_1},\dots,f_{C_m}$ یک Packing است

جلسه بیستم: ۱۴۰۳, ۹, ۱۱

$$\mathcal{N}(\mathcal{F},\|\ \|_\infty,\epsilon)\le\left|C\cap\{X_1,\dots,X_n\}:C\in\mathcal{C}\right|$$

مثال: $\mathcal{C}=\{(-\infty,t],\,t\in\mathbb{R}\}$

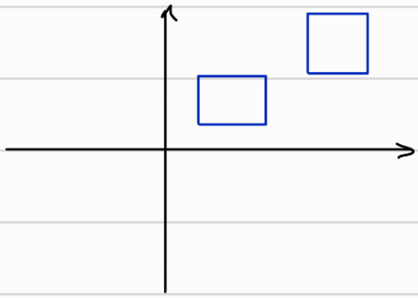
$$(-\infty,t)\cap\{X_1,\dots,X_n\}=\begin{cases}\emptyset & t<X_1\\ \{X_1\} & X_1<t<X_2\\ \{X_1,X_2\} & X_2<t<X_3\\ \vdots & \vdots\\ \{X_1,\dots,X_n\} & t>X_n\end{cases}\Rightarrow |C\cap\{X_1,\dots,X_n\}|=n+1$$

$X_1<\dots<X_n$

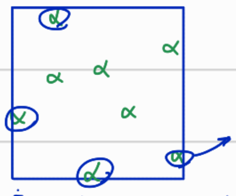
مثال: فرض کنید $\mathcal{C}$ یک مجموعه از همه‌ی زیرمجموعه‌های $\mathbb{R}$ باشد.

$|\mathcal{C}\cap\{X_1,\dots,X_n\}|=2^n \Rightarrow \log\mathcal{N}(\dots)=n\log 2 \rightarrow$ خیلی بد است

مثال: فرض کنید $X_i\in\mathbb{R}^2$ و $\mathcal{C}$ مجموعه‌ی همه‌ی مستطیل‌های موازی با محور مختصات باشد.



$\mathcal{C} \cap \{x_1, \dots, x_n\}$



با تعیین چهارنقطه‌ی بالا و پایین و چپ و راست مستطیل کل نقاط درون آن نیز تعیین می‌شوند.

چون برای هر کدام از نقاط بالا، پایین، چپ و راست n انتخاب وجود دارد، پس $O(n^4)$ اشتراک مختلف وجود دارد. $\Longrightarrow$ $|\mathcal{C} \cap \{x_1, \dots, x_n\}| \leq n^4$

* تعریف بعد VC:

* یک خانواده $\mathcal{C}$ از زیرمجموعه‌های یک فضا مانند $\mathcal{X}$ داده شده است. فرض کنید B یک زیرمجموعه از $\mathcal{X}$ باشد. گوییم $\mathcal{C}$، B را شقه می‌کند اگر

$B \cap \mathcal{C} := \{B \cap C : C \in \mathcal{C}\} =$ همه زیرمجموعه‌های B

* بعد VC یک خانواده $\mathcal{C}$ برابر است با اندازه بزرگترین مجموعه B که توسط $\mathcal{C}$ شقه می‌شود.

$$VC(\mathcal{C}) = \max_{B:\ \text{shatter}} |B|$$

مثال: $\mathcal{C} = \{(-\infty, t) : t \in \mathbb{R}\}$ ، $\mathcal{X} = \mathbb{R}$

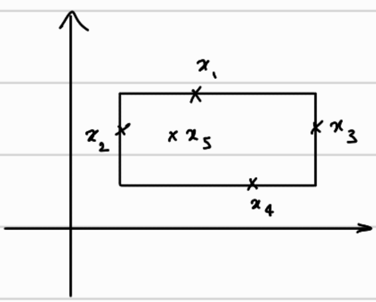
فرض کنید $B = \{a\}$.

$$\left.\begin{array}{l} C = (-\infty, a-1] \Rightarrow B \cap C = \varnothing \\ C = (-\infty, a+1] \Rightarrow B \cap C = \{a\} \end{array}\right\} \Rightarrow B \text{ is Shattered}$$

اگر بگیریم $B = \{a, b\}$ ، $b > a$.

$$(-\infty, t] \cap \{a, b\} = \begin{cases} \varnothing & ; \ t < a \\ \{a\} & ; \ a < t < b \\ \{a, b\} & ; \ t > b \end{cases} \Rightarrow B \text{ can't be Shattered}$$

پس $VC(\mathcal{C}) = 1$.

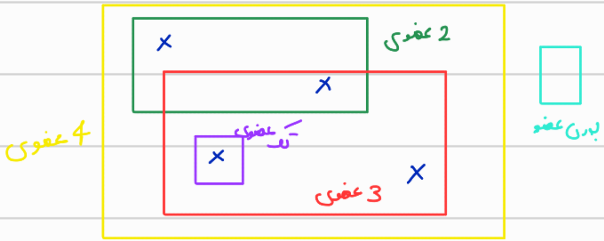


مثال: $\mathcal{C}$ را مجموعه مستطیل‌ها بگیرید. نشان می‌دهیم $VC(\mathcal{C}) = 4$.

ابتدا نشان می‌دهیم $VC < 5$: بگیرید $B = \{x_1, \dots, x_5\}$ مطابق شکل مقابل باشد.

هر طور مستطیل را در نظر بگیریم چون x_5 میان نقاط دیگر است پس $B \cap C \neq \{x_1, \dots, x_4\}$ پس B شقه نمی‌شود.

حال نشان می‌دهیم 4 نقطه شقه می‌شود:



پس $VC(\mathcal{C}) = 4$.

قضیه Sauer-Shelah:

$$|\mathcal{C} \cap \{x_1, \dots, x_n\}| \leq n^{VC}$$

$$|\mathcal{C} \cap \{x_1, \dots, x_n\}| \leq \sum_{k=0}^{VC} \binom{n}{k} \leq \begin{cases} n^{VC} \\ \left(\frac{ne}{VC}\right)^{VC} \end{cases}$$

حال به درس جلسه قبل برمی‌گردیم.

$$E\left[\sup_{c \in \mathcal{C}} \left| P(x \in \mathcal{C}) - \frac{1}{n} \sum 1\{X_i \in \mathcal{C}\} \right| \right] \lesssim \frac{1}{\sqrt{n}} \int_0^1 \sqrt{\log N(\mathcal{F}, \|\ \|_{\mathcal{L}^2(P_n)}, \epsilon)}\, d\epsilon$$

$$\leq \frac{1}{\sqrt{n}} \int_0^1 \sqrt{\log \underbrace{N(\mathcal{F}, \|\ \|_{\mathcal{L}^\infty(P_n)}, \epsilon)}_{\leq |\mathcal{C} \cap \{X_1, \dots, X_n\}| \leq n^{VC}}}\, d\epsilon$$

$$\lesssim \sqrt{\frac{VC(\mathcal{C})}{n} \cdot \log n}$$

چگونه می‌توان $\log n$ را از کران فوق حذف کرد؟

ترفند: به جای استفاده از $\|\ \|_{\mathcal{L}^\infty(P_n)}$ مستقیماً از $\|\ \|_{\mathcal{L}^2(P_n)}$ استفاده کنیم.

نشان خواهیم داد: $N(\mathcal{F}, \|\ \|_{\mathcal{L}^2(q)}, \epsilon) \leq r^{VC}$: توزیع دلخواه q $\forall$ که در آن $r = \frac{k}{\epsilon^4}$ ، k ثابت است.

شهود اثبات:

$$N(\mathcal{F}, \|\ \|_{L^2(q)}, \epsilon) \leq P(\mathcal{F}, \|\ \|_{L^2(q)}, \epsilon)$$

اگر $P(\dots) = L$ باشد یعنی $f_1, \dots, f_L$ وجود دارند که $E_q\left[(f_i(x) - f_j(x))^2\right] \geq \epsilon^2$.

فرض کنید $X_1,\dots,X_r$ ، r نمونه iid از توزیع q باشند.

$$\frac{1}{r}\sum_{k=1}^{r}(f_i(X_k)-f_j(X_k))^2 \approx E_q[(f_i-f_j)^2]$$

$$\|f_i-f_j\|_{L^2(P_r)} \approx E_q[(f_i-f_j)^2] \Rightarrow$$ انتظار داریم $\|f_i-f_j\|_{L^2(P_r)} \geq (\frac{\epsilon}{2})^2$ یعنی فاصله در این معیار نیز بزرگ باشد

$\|f_i-f_j\|_{L^\infty(P_r)}=1 \Rightarrow$ روی r نقطه از $X_1,\dots,X_r$; $f_i\neq f_j \Rightarrow$

نشان می‌دهیم $r\simeq\frac{k}{\epsilon^4} \to L\leq N(F,\|\,\|_{L^2(P_r)},\epsilon)\leq r^{VC} \Rightarrow$

اثبات کامل در جلسه آینده ارائه می‌شود.

جلسه بیست و یکم: 13 / 9 / 1403

* یادآوری:

$$\underset{X_1,\dots,X_n\sim\mu}{E}\left[\sup_{f\in F} E_\mu(f(x))-\frac{1}{n}\sum_{i=1}^{n}f(x_i)\right]\leq\sqrt{\frac{\log n.\,VC(F)}{n}}$$

* کران به μ ربطی ندارد.
* با دانستن μ ممکن است کران های بهتری بتوان یافت.

در جلسات قبلی:

$$E[\sup\dots]\lesssim\frac{1}{\sqrt{n}}\int_0^1\sqrt{\log N(F,L^2(P_n),\epsilon)}\,d\epsilon$$

$$\leq\frac{1}{\sqrt{n}}\int_0^1\sqrt{\underbrace{\log N(F,L^\infty(P_n),\epsilon)}_{\leq VC.\log n}}\,d\epsilon \to$$ نقطه‌ای که باعث ایجاد $\log n$ شد

لم: فرض کنید $\{f_1,\dots,f_D\}$ یک مجموعه ϵ-بسته بندی نسبت به متر $L^2(q)$ (توزیع دلخواه q) باشد یعنی

$$E_q[(f_i(X)-f_j(X))^2]\geq\epsilon^2 \quad ; \quad \|f\|_\infty=1$$

آنگاه برای $r\geq\frac{k}{\epsilon^4}\log D$ می‌توان نقاط $x_1,\dots,x_r$ پیدا کرد به طوری که

$$\frac{\epsilon^2}{4}\leq E_{P_r}[(f_i(x)-f_j(x))^2]=\frac{1}{r}\sum_{k=1}^{r}(f_i(x_k)-f_j(x_k))^2$$

با فرض درست بودن لم بالا خواهیم داشت : $\mathcal{N}(\mathcal{F}, \mathcal{L}^2(P_n), \epsilon) \leq \overbrace{P(F, \mathcal{L}^2(P_n), \epsilon)}^{D} \overset{\text{طبق لم بالا}}{\leq} P(\mathcal{F}, \mathcal{L}^2(P_r), \frac{\epsilon}{2})$

دقت کنید که روی x_r ، ... ، x_1 ، حداقل یک k وجود دارد که $f_i(x_k) \neq f_j(x_k)$ (چون فاصله توابع از $\frac{\epsilon}{2}$ بزرگتر است). حال در مسئله اصلی دقت کنید f_iها توابع

مشخصه ($\mathbb{1}[x \leq t]$) هستند. پس $f_i(x) = \mathbb{1}[x \in A_i] \; ; \; A_i \in \mathcal{C}$

چون $f_i(x_k) \neq f_j(x_k)$ است پس $\exists x_k : x_k \in A_i \Delta A_j \Rightarrow A_i \cap \{x_1, \dots, x_r\} \neq A_j \cap \{x_1, \dots, x_r\}$

$$\overset{\text{چون D تا } A_i \text{ وجود دارد}}{\Longrightarrow} |\mathcal{C} \cap \{x_1, \dots, x_r\}| \geq D$$

همچنین جلسه قبل دیدیم $|\mathcal{C} \cap \{x_1, \dots, x_r\}| \leq \left(\frac{er}{vc}\right)^{vc}$. پس $D \leq \left(\frac{e}{vc} \cdot \frac{k}{\epsilon^4} \log D\right)^{vc}$

$$\log D = \frac{1}{\alpha} \log D^\alpha \leq \frac{1}{\alpha} D^\alpha$$

$$\Rightarrow D \leq \left(\tilde{k} \frac{1}{2vc} \frac{\log D}{\epsilon^4}\right)^{vc} \leq \left(\tilde{k} \cdot \frac{D^{\frac{1}{2vc}}}{\epsilon^4}\right)^{vc} = \left(\frac{\tilde{k}}{\epsilon^4}\right)^{vc} \sqrt{D} \Rightarrow D \leq \left(\frac{\tilde{k}}{\epsilon^4}\right)^{2vc} = \left(\frac{c}{\epsilon}\right)^{8vc}$$

$$\Rightarrow \log \mathcal{N}(\mathcal{F}, \mathcal{L}^2(P_n), \epsilon) \leq \log D \leq 8vc \log \frac{c}{\epsilon}$$

$$\Rightarrow E[\sup \dots] \lesssim \frac{1}{\sqrt{n}} \int_0^1 \sqrt{\log \mathcal{N}(\dots)} \lesssim \frac{\sqrt{vc}}{\sqrt{n}} \int_0^1 \sqrt{\underbrace{\log \frac{c}{\epsilon}}_{< \frac{c}{\epsilon}}} \, d\epsilon \lesssim \boxed{\sqrt{\frac{vc(\mathcal{C})}{n}}}$$

حال به اثبات لم می‌پردازیم. این اثبات وجودی است یعنی نشان می‌دهیم اگر $X_i \sim q$ به شکل iid انتخاب شوند

$Pr[\forall i,j : \|f_i(X) - f_j(X)\|_{\mathcal{L}^2(q_r)} > \frac{\epsilon}{2}] > 0 \Rightarrow$ یک تحقق $\{x_1, \dots, x_r\}$ وجود دارد.

بر جای احتمال بالا متمم آن را بررسی می‌کنیم یعنی

$$Pr\left[\exists i,j : \frac{1}{r} \sum (f_i(X_k) - f_j(X_k))^2 < \frac{\epsilon^2}{4}\right] \leq \sum_{(i,j)} Pr\left[\frac{1}{r} \sum (f_i(X_k) - f_j(X_k))^2 < \frac{\epsilon^2}{4}\right]$$

فرض : $\|f_i(X) - f_j(X)\|_{L^2(q)} > \epsilon$ یا به عبارت دیگر $E_q[(f_i(X) - f_j(X))^2] > \epsilon^2$.

$$Pr\left[\frac{1}{r}\sum (f_i(x_k)-f_j(x_k))^2 < \frac{\epsilon^2}{4}\right] \le Pr\left[\underbrace{E(f_i-f_j)^2}_{\geq \epsilon^2} - \overbrace{\frac{1}{r}\sum \underline{(f_i(x_k)-f_j(x_k))^2}}^{\leq \epsilon^2/4} > \frac{3\epsilon^2}{4}\right]$$

$Z_k \le 4$: $\|f_i\| \le 1$ چون فرض کردیم

یعنی $-1 \le f_i \le 1$

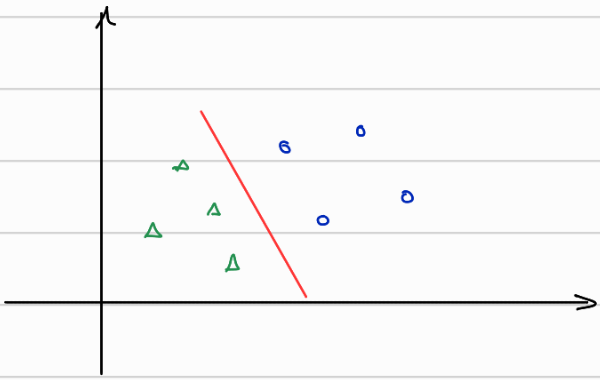
هوفدینگ

$$\le \exp\left(-c\cdot\left(\frac{3\epsilon^2}{4}\right)^2\cdot r\right) = \exp(-c'\cdot r\cdot \epsilon^4)$$

تعداد زوج‌های (زوج)

$$\Rightarrow Pr\left[\exists i,j : \frac{1}{r}\sum (f_i(x_k)-f_j(x_k))^2 < \frac{\epsilon^2}{4}\right] \le D^2 \exp(-c'\cdot r\cdot\epsilon^4) < 1 \Rightarrow r \geq c''\frac{\log D}{\epsilon^4}$$

* کاربرد:



واقعیت: $(\vec{X}, y)$ ، $y \in \{0,1\}$. یک توزیع مشترک $P_{\vec{x},y}$ وجود داشته باشد.

به توزیع $P_{\vec{x},y}$ فقط از طریق n نمونه‌ی $\{(X_1,y_1), \dots, (X_n,y_n)\}$ دسترسی داریم.

هدف: پیدا کردن تابعی مانند $f \in \mathcal{F}$ به طوری که $\underset{X,Y}{Pr}[f(x) \neq y]$ روی خانواده‌ی $\mathcal{F}$ کمینه شود.

تعریف: خطای واقعی یک تابع مانند f : $R(f) = Pr[f(x) \neq y]$, $f^* = \underset{f \in \mathcal{F}}{\arg\min}\ R(f)$

ریسک تجربی: $R_n(f) = \frac{1}{n}\sum 1\{f(X_i) \neq y_i\}$

ERM: $f_n^* = \underset{f \in \mathcal{F}}{\arg\min}\ R_n(f)$

* $R(f^*)$ کمترین خطای ممکن است. خطای الگوریتم ERM برابر با $R(f_n^*)$

* خطای افزوده (Excess Risk): $R(f_n^*) - R(f^*)$

$Pr[f(x) \neq y]$

فرض کنید $\epsilon = \sup_{f \in \mathcal{F}} |R(f) - R_n(f)|$ ← تابع داده آموزش

$\frac{1}{n}\sum 1[f(x) \neq y]$

لم: $0 \leq R(f_n^*) - R(f^*) \leq 2\epsilon$

اثبات: $R(f_n^*) - R(f^*) \leq R_n(f_n^*) - R(f^*) + \epsilon$

$$\leq R_n(f^*) - R(f^*) + \epsilon \leq 2\epsilon$$

جلسه بیست و دوم: 1403_9_18

هدف: باندزدن روی $E\left[\,|R(f_n^*) - R(f^*)|\,\right]$. طبق جلسه قبل:

$$E\left[\,|R(f_n^*) - R(f^*)|\,\right] \leq 2 \underset{\substack{X_1,\dots,X_n\\ Y_1,\dots,Y_n}}{E}\left[\sup_{f\in F} |R(f) - R_n(f)|\right]$$

$$= 2 \underset{P_{XY}^{\otimes n}}{E}\left[\sup_{f\in F}\left| E\big[\underbrace{1\{f(X)\neq Y\}}_{g(X,Y)}\big] - \frac{1}{n}\sum \underbrace{1\{f(X_i)\neq Y_i\}}_{g(X_i,Y_i)}\right|\right]$$

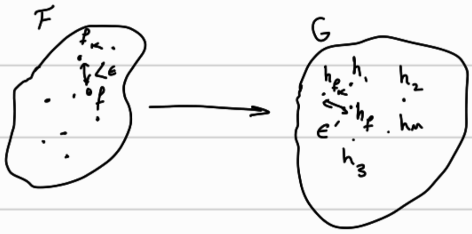
$$= 2E\left[\sup_{g\in G}\left| E[g(X,Y)] - \frac{1}{n}\sum g(X_i,Y_i)\right|\right] \lesssim \sqrt{\frac{VC(G)}{n}}$$

$$G = \{\overbrace{1\{f(X)\neq Y\}}^{h_f(X,Y)} : f\in F\} \longrightarrow$$

به دنبال جایگزینی $VC(G)$ با $VC(F)$ $\Leftarrow$ بدست آوردن VC برای این کلاس توابع سخت است
هستیم که قابل محاسبه است.

$$E\left[\sup_{g\in G} E[g(X,Y)] - \frac{1}{n}\sum g(X_i,Y_i)\right] \lesssim \int_0^1 \sqrt{\log N(G, \mathcal{L}^2(P_n), \epsilon)}\, d\epsilon$$

هدف ثانویه: رابطه‌ی $N(G,\mathcal{L}^2(P_n),\epsilon)$ با $N(F,\mathcal{L}^2(P_n),\epsilon)$. فرض کنید یک پوشش مانند $\{f_1,\dots,f_m\}$ از F داده شده باشد.



$$\Longrightarrow N(G,\epsilon') \leq N(F,\epsilon)$$

$$\|f - f_k\|^2_{\mathcal{L}^2(P_n)} = E\left[(f(x) - f_k(x))^2\right] = E\left[(1\{f(x)\neq f_k(x)\})\right] = \Pr\left[f(x)\neq f_k(x)\right]$$

$$\|\underset{h_f}{h} - \underset{h_{f_k}}{h_k}\|^2 = E\left[1\{h_f(x)\neq h_{f_k}(x)\}\right] = \Pr\left[h_f(x)\neq h_{f_k}(x)\right]$$

$$= \Pr\left\{ \mathbb{1}\{f(X) \neq Y\} \neq \mathbb{1}\{f_k(X) \neq Y\} \right\}$$

$$= \Pr\left\{ f(X) \neq f_k(X) \right\} \rightarrow$$ فرض شده است که $Y \in [0,1]$

$$\Rightarrow \|f - f_k\|_{\mathcal{L}^2(P_n)} = \|h_f - h_{f_k}\|_{\mathcal{L}^2(P_n)} \Rightarrow \mathcal{N}(G, \mathcal{L}^2(P_n), \epsilon) \leq \mathcal{N}(F, \mathcal{L}^2(P_n), \epsilon)$$

$$\Rightarrow 2E[\epsilon] \lesssim \int_0^1 \sqrt{\log \mathcal{N}(G, \mathcal{L}^2(P_n), \epsilon)}\, d\epsilon \leq \int_0^1 \sqrt{\log \mathcal{N}(F, \mathcal{L}^2(P_n), \epsilon)}\, d\epsilon \lesssim \sqrt{\frac{VC(F)}{n}}$$

- ماتریس های تصادفی:

رفرنس ها:

- فصل 6 ورژن
- An Introduction to Matrix Concentration Inequalities; Joel Tropp

یادآوری: محاسبه‌ی تابع از یک ماتریس: (همیشه با ماتریس‌های متقارن کار می‌کنیم)

راه اول: $e^A = I + A + \frac{A^2}{2!} + \dots$

راه دوم: $A = UDU^T \;;\; e^A = U \begin{pmatrix} e^{d_1} & & 0 \\ & \ddots & \\ 0 & & e^{d_k} \end{pmatrix} U^T$

فرض کنید $f: \mathbb{R} \to \mathbb{R}$. $A = \sum_{i=1}^n \lambda_i u_i u_i^T \Leftarrow$ $f: S_n \to S_n :\quad f(A) = \sum_{i=1}^n f(\lambda_i) u_i u_i^T$

مثال: $|A| = \sum_{i=1}^n |\lambda_i| u_i u_i^T$

- مفهوم یکنوا یی: $A \leq B \Rightarrow$ $B - A$ مثبت نیمه معین باشد

فرض کنید $f: \mathbb{R} \to \mathbb{R}$ تابعی صعودی باشد. آیا $A \leq B \Rightarrow f(A) \leq f(B)$؟ پاسخ خیر است. اگر فرض کنید $f(x) = x^2$ با دامنه‌ی $x \geq 0$ باشد. می‌توان ماتریس های A و B را طوری پیدا کرد که $A \leq B$ اما $A^2 \leq B^2$.

- $f(t) = \frac{-1}{t}$ یکنواست. $A \leq B \overset{?}{\Longrightarrow} B^{-1} \leq A^{-1}$