دانشکده‌ی مهندسی برق

آنالیز احتمالاتی در ابعاد بالا	بهار ۱۴۰۲
تمرین سری دوم	
مدرس: دکتر محمد حسین یاسائی	موعد تحویل: ۲۷ فروردین

حل سوالات ۱ تا ۳ اجباری است. از بین سوال‌های ۴ تا ۹ چهار سوال را به دلخواه حل کنید. سوال ۱۰ و ۱۱ نیز امتیازی هستند.

۱ عدد پوششی

سه بخش این سوال ارتباطی با یکدیگر ندارند.

۱.۱ عدد پوششی خارجی

فرض کنید مجموعه‌ی A و متر d روی اعضای A داده‌شده باشد. مجموعه‌ی K را یک زیرمجموعه از A در نظر بگیرید. در کلاس درس، با مفهوم عدد پوششی آشنا شدیم.

- در یک تعریف عدد پوششی، فرض بر این است که مراکز گوی‌های پوشش‌دهنده‌ی مجموعه‌ی K، یعنی نقاط x_i، همگی اعضای مجموعه‌ی K باشند. عدد پوششی به دست آمده از این تعریف را $\mathcal{N}(K, d, \epsilon)$ می‌نامیم که ϵ شعاع گوی‌های پوشش‌دهنده است.
- در تعریف دوم، مراکز گوی‌های پوشش دهنده، الزامی به عضو K بودن ندارند و می‌توانند هر عضو دلخواه A باشند. عدد پوششی به دست آمده از این تعریف را $\mathcal{N}^{ext}(K, d, \epsilon)$ می‌نامیم.

به توجه به این دو تعریف، نشان دهید:

$$\mathcal{N}^{ext}(K, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon) \leq \mathcal{N}^{ext}(K, d, \epsilon/2)$$

۲.۱ یکنوایی عدد پوششی

در ابتدا، مثال نقضی برای نامساوی زیر بزنید

$$L \subset K \implies \mathcal{N}(L, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon)$$

سپس تلاش نمایید نامساوی زیر را ثابت کنید

$$L \subset K \implies \mathcal{N}(L, d, \epsilon) \leq \mathcal{N}(K, d, \epsilon/2).$$

۳.۱ پوشش و گنجایش مکعب همینگ

مجموعه‌ی $K = \{0, 1\}^n$ را با فاصله‌ی همینگ در نظر بگیرید. نشان دهید که برای هر عدد حسابی $m \leq n$ داریم:

$$\frac{2^n}{\sum_{k=1}^{m} \binom{n}{k}} \leq \mathcal{N}(K, d_H, m) \leq \mathcal{M}(K, d_H, m) \leq \frac{2^n}{\sum_{k=0}^{[\frac{m}{2}]} \binom{n}{k}}$$

که در آن d_H متر همینگ، $\mathcal{N}$ عدد پوششی و $\mathcal{M}$ عدد گنجایشی است.

۴.۱ ابرمکعب باینری

$\mathbb{H}^d = \{0,1\}^d$ و متر همینگ[1] اسکیل شده $d_H(a,a') = \frac{1}{d}\sum_{j=1}^{d} \mathbb{I}(a_j \neq a'_j)$ را درنظر بگیرید. ثابت کنید عدد گنجایشی[2] این فضای متری دارای کران بالای زیر است.

$$\frac{\log(\mathcal{M}(\mathbb{H}^d, d_H, \delta))}{d} \leq D(\delta/2\|1/2) + \frac{\log(d+1)}{d}$$

که $D(\delta/2\|1/2) = \frac{\delta}{2}\log(\frac{\delta/2}{1/2}) + (1-\frac{\delta}{2})\log(\frac{1-\delta/2}{1/2})$ همان آنتروپی نسبی باینری[3] است.

۲ کران پایین برای ماکزیمم متغیرهای تصادفی

در این مساله می‌خواهیم نشان دهیم کران بالایی که برای امیدریاضی ماکزیمم تعدادی متغیر تصادفی به دست آورده‌ایم در حالتی که متغیرها از هم مستقل باشند، کران بسیار خوبی است و از همان مرتبه می‌توان کران پایینی برای امیدریاضی ماکزیمم متغیرهای تصادفی به دست آورد. برای این کار گام‌های زیر را طی می‌کنیم.

۱. اگر $A_1, \ldots, A_n$ پیشامدهای مستقلی باشند، نشان دهید

$$\left(1-e^{-1}\right)\left\{1 \wedge \sum_{k=1}^{n} \mathbf{P}\left[A_k\right]\right\} \leq \mathbf{P}\left[\bigcup_{k=1}^{n} A_k\right] \leq 1 \wedge \sum_{k=1}^{n} \mathbf{P}\left[A_k\right]$$

که منظور از $a \wedge b$ همان $\min(a,b)$ است.

راهنمایی: دقت کنید که $\prod_{k=1}^{n}\{1-x_k\} \leq \exp\left(-\sum_{k=1}^{n} x_k\right)$ و $1 - e^{-x} \geq \left(1-e^{-1}\right) 1 \wedge x$.

۲. فرض کنید η^* یک تابع اکیداً صعودی و محدب باشد.هم‌چنین فرض کنید که

$$\mathbf{P}\left[X_t \geq x\right] \geq e^{-\eta^*(x)} \quad \forall x \geq 0 \;\;,\;\; \forall t \in T$$

و $\{X_t : t \in T\}$ از یکدیگر مستقل هستند. نشان دهید برای هر $u \geq 0$ داریم

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq \eta^{*-1}(\log|T| + u)\right] \geq \left(1-e^{-1}\right) e^{-u}$$

این کران را با کرانی که در قسمت دوم سوال نامساوی ماکسیمال به دست آورده‌اید مقایسه کنید.

حال که ما یک کران پایین روی احتمال دُم ماکزیمم تعدادی متغیر تصادفی به دست آوردیم، می‌توانیم یک کران پایین روی امیدریاضی آن نیز به کمک انتگرال‌گیری روی احتمال دم به دست آوریم.

۳. از قسمت قبل نتیجه بگیرید که برای هر $x \geq 0$ داریم

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq \eta^{*-1}(2\log|T|)/2 + x\right] \geq \left(1-e^{-1}\right) e^{-\eta^*(2x)/2}.$$

راهنمایی: از مقعر بودن η^{*-1} استفاده کنید.

۴. فرض کنید ψ^* نیز یک تابع اکیداً صعودی و محدب باشد. هم‌چنین فرض کنید که

$$e^{-\eta^*(x)} \leq \mathbf{P}\left[X_t \geq x\right] \leq e^{-\psi^*(x)} \quad \forall x \geq 0 \;\;,\;\; \forall t \in T$$

حال نتیجه بگیرید که ثابت‌های مثبت C_1 و C_2 وجود دارند به نحوی که

$$C_1\left(\eta^{*-1}(\log|T|) + \sup_{t\in T}\mathbf{E}\left[0 \wedge X_t\right]\right) \leq \mathbf{E}\left[\sup_{t\in T} X_t\right] \leq C_2\left(\psi^{*-1}(\log|T|)\right)$$

راهنمایی: از $\mathbf{E}[\max(0,Z)] = \int_0^\infty \mathbf{P}[Z \geq x]dx$ استفاده کنید.

کران‌های بالا و پایین که در این قسمت به دست آورده‌ایم معمولا از یک مرتبه هستند به شرط اینکه کران‌های بالا و پایینی که در شروع روی $\mathbf{P}\left[X_t \geq x\right]$ می‌گذاریم از یک مرتبه باشند. به عنوان مثال متغیرهای گوسی را بررسی می‌کنیم.

[1] distance hamming
[2] packing
[3] entropy relative binary

۵. برای متغیر تصادفی $X \sim \mathcal{N}(0,1)$ نشان دهید

$$\mathbf{P}[X \geq x] \geq \frac{e^{-x^2}}{2\sqrt{2}} \quad \forall x \geq 0$$

راهنمایی: احتمال را به صورت انتگرالی نوشته و از نامساوی $(v+x)^2 \leq 2v^2 + 2x^2$ بهره ببرید.

۶. فرض کنید $X_1, \dots, X_n$ متغیرهای تصادفی مستقل و هم‌توزیع با توزیع نرمال استاندارد باشند. از قسمت‌های قبل کمک گرفته و نشان دهید

$$\frac{1-e^{-1}}{2}\sqrt{2\log n 2^{-3/4}} - \frac{1}{\sqrt{2\pi}} \leq \mathbf{E}\left[\max_{i\leq n} X_i\right] \leq \sqrt{2\log n}.$$

و به صورت خاص، برای nهای به حد کافی بزرگ داریم $c\sqrt{\log n} \leq \mathbf{E}\left[\max_{i\leq n} X_i\right] \leq C\sqrt{\log n}$.

۳ مساله‌ی فروشنده‌ی دوره‌گرد

فرض کنید $X_1, \dots, X_n$ نقاطی تصادفی و i.i.d. باشند که به صورت یکنواخت روی مربع واحد $[0,1]^2$ توزیع شده‌اند. ما به X_i به چشم مکان شهر iاُم در مربع واحد نگاه می‌کنیم. هدف مساله‌ی فروشنده‌ی دوره‌گرد پیدا کردن یک گشتی است که یک بار از همه‌ی شهرها بگذرد و کم‌ترین مسافت ممکن را طی کرده باشد. ما طول کم‌ترین گشت را به صورت

$$L_n := \min_{\sigma}\left\{\left\|X_{\sigma(1)} - X_{\sigma(2)}\right\| + \left\|X_{\sigma(2)} - X_{\sigma(3)}\right\| + \cdots + \left\|X_{\sigma(n)} - X_{\sigma(1)}\right\|\right\}$$

نشان می‌دهیم، که در آن کمینه‌سازی روی همه‌ی جایگشت‌های ممکن از $\{1, \dots, n\}$ است. برای شروع به میزان بزرگی L_n می‌پردازیم.

۱. نشان دهید $\mathbf{E}\left[L_n\right] \asymp \sqrt{n}$.

راهنمایی: برای کران پایین ثابت کنید که $L_n \geq \sum_{k=1}^{n} \min_{l\neq k} \|X_k - X_l\|$ و برای کران بالا $L_n \leq L_{n-1} + 2\min_{k<n} \|X_n - X_k\|$.

۲. از نامساوی McDiarmid استفاده کنید که نشان دهید L_n یک متغیر تصادفی $2n-$زیرگوسی است.

کرانی که با استفاده از نامساوی McDiarmid به دست می‌آید افتضاح است! این کران نتیجه می‌دهد که بزرگی تغییرات و نوسانات حول میانگین هم‌مرتبه از خود آن است. در نتیجه نامساوی McDiarmid حتی نتیجه نمی‌دهد که L_n حول متوسط‌اش متمرکز است. اینجاست که استفاده از نامساوی Talagrand نتایج بسیار بهتری حاصل می‌کند و به کمک آن می‌توان تمرکز اندازه حول میانگین برای L_n را نشان داد. نامساوی Talagrand در کلاس بیان و اثبات نشده است. ما این نامساوی را بدون اثبات در اینجا می‌آوریم و این کاربرد جالب از آن را می‌بینیم.

Talagrand's inequality: فرض کنید $X_1, \dots, X_n$ مستقل باشند و داشته باشیم

$$f(x) - f(y) \leq \sum_{i=1}^{n} c_i(x)\mathbf{1}_{x_i\neq y_i} \quad \forall x, y \in \mathbb{R}^n.$$

در نتیجه $f\left(X_1, \dots, X_n\right)$ یک متغیر $\left\|\sum_{i=1}^{n} c_i^2\right\|_\infty$-زیرگوسی است.

پیش از استفاده از این نامساوی، کمی بینش هندسی عمیق‌تری نسبت به مساله نیاز داریم که در چند قسمت ابتدایی بعدی به آن می‌پردازیم.

۳. فرض کنید $v = (0, a)$ و $w = (b, 0)$ رئوس مثلث قائم‌الزاویه‌ی $T = \operatorname{conv}\{0, v, w\}$ باشند. نشان دهید

$$\forall x \in T \quad \|v - x\|^2 + \|x - w\|^2 \leq \|v - w\|^2$$

۴. اثبات کنید که برای هر $x_1, \dots, x_n \in T$ یک جایگشت مانند σ از $\{1, 2, \dots, n\}$ وجود دارد به نحوی که

$$\left\|v - x_{\sigma(1)}\right\|^2 + \sum_{i=1}^{n-1}\left\|x_{\sigma(i)} - x_{\sigma(i+1)}\right\|^2 + \left\|x_{\sigma(n)} - w\right\|^2 \leq \|v - w\|^2$$

راهنمایی: از استقرا استفاده کنید. فرض کنید حکم برای هر مثلث قائم‌الزاویه مانند S و $x_1, \dots, x_{n-1} \in S$ برقرار است. مثلث T را با رسم ارتفاع از مبدا به وتر به دو مثلث قائم‌الزاویه‌ی کوچک‌تر تقسیم کنید. اگر هر کدارم از زیر مثلث‌ها شامل نقاطی بودند از فرض استقرا استفاده کنید تا حکم را نتیجه بگیرید. در غیر این‌صورت اینقدر به تقسیم کردن مثلث‌ها ادامه دهید تا بتوان از فرض استقرا استفاده کرد.

۵. حال نتیجه بگیرید که برای هر مجموعه نقاطی مانند $x_1, \dots, x_n \in [0,1]^2$ یک جایگشت مانند σ از $\{1, 2, \dots, n\}$ وجود دارد به نحوی که $\|x_{\sigma(1)} - x_{\sigma(2)}\|^2 + \|x_{\sigma(2)} - x_{\sigma(3)}\|^2 + \cdots + \|x_{\sigma(n)} - x_{\sigma(1)}\|^2 \leq 4$.

حال ما از این بینش عمیق‌تر هندسی خود استفاده می‌کنیم تا طول گشت فروشنده‌ی دوره‌گرد را تحلیل کنیم. به یاد داریم که یک گشت بین نقاط $x_1, \dots, x_n$ به کمک یک جایگشت σ از $\{1, 2, \dots, n\}$ تعریف می‌شد. طول یک گشت مشخص را با نماد $l_n(x, \sigma)$ نشان می‌دهیم. در نتیجه $L_n := \min_\sigma l_n(X, \sigma)$.

۶. فرض کنید $x = \{x_1, \dots, x_n\}$ و $y = \{y_1, \dots, y_n\}$ دو مجموعه نقاطی در $[0,1]^2$ باشند که $x \cap y \neq \varnothing$. فرض کنید σ یک گشت برای x و τ یک گشت برای y باشد. نشان دهید یک گشت مانند ρ از $x \cup y$ وجود دارد به نحوی که

$$l_{2n}(x \cup y, \rho) \leq l_n(y, \tau) + 2\sum_{i=1}^{n} \mathbf{1}_{x_i \notin y} d_i(x, \sigma)$$

که در آن $d_i(x, \sigma)$ فاصله‌ی بین x_i و نقطه‌ی قبلی در گشت σ است.

راهنمایی: فرض کنید σ و τ دو مسیر پیاده روی هستند که با رنگ‌های قرمز و آبی مشخص شده اند و در بعضی شهرها با هم تلاقی دارند. هدف شما این است که به طور سیستماتیک اجتماع مسیرها را پیاده‌روی کنید. برای این منظور، پیاده روی زیر را انجام دهید: شروع به راه رفتن در مسیر آبی کنید. اگر در نقطه‌ای مسیر قرمز از مسیر آبی جدا می‌شد، مسیر قرمز را تا زمانی که دوباره به مسیر آبی برخورد کند، طی کنید، سپس به جایی که از مسیر آبی منحرف شده‌اید، برگردید و مسیر آبی را ادامه دهید. در حالی که این پیاده روی یک گشت نیست (چون برخی از نقاط دو بار بازدید می شوند)، می‌توانید بدون افزایش طول آن، آن را به یک گشت واقعی تبدیل کنید.

۷. برای هر $x_1, \dots, x_n \in [0,1]^2$ گشت σ_x را همان گشتی در نظر بگیرید که خاصیت قسمت ۵ را دارد. نشان دهید برای هر $x, y \in [0,1]^{2n}$ داریم $\min_\sigma l_n(x, \sigma) \leq \min_\sigma l_n(y, \sigma) + \sum_{i=1}^{n} 2d_i\left(x, \sigma_x\right) \mathbf{1}_{x_i \neq y_i}$.

۸. نتیجه بگیرید که L_n برای هر $n \geq 1$ یک متغیر 16-زیرگوسی می‌باشد.

۴ نرم ماتریس زیرگوسی

فرض کنید A یک ماتریس تصادفی $m \times n$ با ردیف‌های A_i باشد که A_i ها بردارهای مستقل، با میانگین صفر، زیرگوسی و ایزوتروپیک هستند. نشان دهید که برای هر t مثبت، عبارت زیر با احتمال حداقل $1 - 2\exp(-t^2)$ برقرار است:

$$\left|\left|\frac{1}{m}A^T A - I_n\right|\right|_{op} \leq C \max(\delta, \delta^2)$$

که در آن $\delta = \sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}$ است.

یادداشت: بردار تصادفی X زیرگوسی است اگر به ازای هر بردار x عضو کره‌ی واحد، ضرب داخلی X و x زیرگوسی با ثابت محدود باشد. در قضیه‌ی فوق، ثابت‌های C می‌توانند به ضرایب زیرگوسی بودن بردار X نیز وابسته باشند.

راهنمایی: نرم اپراتوری را به فرم سوپریمی آن نوشته و تلاش کنید با انتخاب یک ϵ-net مناسب از کره‌ی واحد و با ایده‌ی گسسته‌سازی، کرانی برای احتمال مطلوب بیابید. از نامساوی برن‌اشتاین بهره ببرید. توجه کنید که تنها یک مرحله گسسته‌سازی کافی است و نیازی به استفاده از ایده‌های چینینگ نیست.

۵ کران روی واریانس ماکسیمم گاوسی‌ها

فرض کنید $X \sim \mathcal{N}(0, \Sigma)$ یک بردار تصادفی گوسی n-بعدی با میانگین صفر و ماتریس کوواریانس دلخواه Σ باشد. نامساوی زیر را اثبات کنید:

$$\mathrm{Var}\left[\max_{i=1,2,\dots,n} X_i\right] \leq \max_{i=1,2,\dots,n} \mathrm{Var}\left[X_i\right]$$

راهنمایی: بردار تصادفی X را به صورت $X = \Sigma^{\frac{1}{2}} Y$ بنویسید که $Y_1, Y_2, \dots, Y_n$ متغیرهای تصادفی گوسی استاندارد و مستقل از هم هستند.

۶ کران واریانس برای متغیر تصادفی پواسون

۱. متغیر تصادفی دوجمله‌ای با پارامترهای n و p را درنظر بگیرید، یعنی $\mathbb{P}[B=k]=\binom{n}{k}p^k(1-p)^{n-k}$. برای هر تابع دلخواه $f:\{0,\ldots,n\}\to\mathbb{R}$ نشان دهید:

$$\begin{aligned} Var(f(B)) &\leq p(1-p)\mathbb{E}\left[B(f(B)-f(B-1))^2+(n-B)(f(B+1)-f(B))^2\right] \\ &= p\mathbb{E}[(n-B)(f(B+1)-(B))^2] \end{aligned}$$

۲. فرض کنید $p=\frac{\mu}{n}$. در این حالت متغیر تصادفی B برای nهای بزرگ به متغیر تصادفی پواسون X با میانگین μ میل می‌کند. نشان دهید اگر $\sup_k|f(k+1)-f(k)|<\infty$،

$$\operatorname{Var}(f(X))\leq\mu\mathbb{E}\left[(f(X+1)f(X))^2\right]$$

۳. برای متغیر تصادفی پواسون X با میانگین μ در بسیاری از مسائل آمار کاربردی از تبدیل تثبیت‌کننده مربع $Y=\sqrt{X}$ در جهت کاهش وابستگی میزان واریانس یا انحراف معیار متغیر جدید به میانگین، استفاده می‌کنند. نشان دهید

$$\operatorname{Var}(Y)\leq\mu\mathbb{E}\left[\frac{1}{4X+1}\right]$$

۷ بسته‌بندی بارهای کشتی!

به عنوان راهنمایی در شروع می‌گوییم که این سوال یکی از کاربردهای قدیمی و معروف نامساوی افرون-اشتاین [4] است.

۱. برای شروع نتیجه‌ی مهم زیر از نامساوی افرون-اشتاین را ثابت کنید.

$$\operatorname{Var}\left[f\left(X_1,\ldots,X_n\right)\right]\leq\frac{1}{4}\mathbf{E}\left[\sum_{i=1}^n\left(D_if\left(X_1,\ldots,X_n\right)\right)^2\right]$$

که در آن

$$D_if(x):=\sup_z f\left(x_1,\ldots,x_{i-1},z,x_{i+1},\ldots,x_n\right)-\inf_z f\left(x_1,\ldots,x_{i-1},z,x_{i+1},\ldots,x_n\right)$$

۲. حال فرض کنید $X_1,X_2,\ldots,X_n$ متغیرهای تصادفی i.i.d. با مقادیر در بازه‌ی $[0,1]$ با امیدریاضی $\frac{1}{2}$ باشند. هر کدام از این متغیرهای تصادفی نشان‌دهنده‌ی اندازه‌ی میزان باری است که باید توسط کشتی جابه‌جا شود. هر کدام از این بارها باید در محفظه‌هایی به اندازه‌ی ۱ جاساز شوند تا بتوانند داخل کشتی انبار شده و سپس حمل شوند. بنابراین هر محفظه می‌تواند مجموعه‌ای از بارها که جمع اندازه‌شان حداکثر ۱ است را در داخل خود جا دهد. تعریف می‌کنیم $B_n=f(X_1,X_2,\ldots,X_n)$ کم‌ترین تعداد محفظه‌هایی باشد که بارها همگی در آن‌ها جاساز شوند. جالب است بدانید محاسبه‌ی B_n یک مساله‌ی بهینه‌سازی ترکیبیاتی سخت به حساب می‌آید ولی می‌توان میانگین و واریانس آن را به سادگی کران زد.

(آ) نشان دهید $\operatorname{Var}[B_n]\leq\frac{n}{4}$.

(ب) نشان دهید $\mathbf{E}[B_n]\geq\frac{n}{2}$.

از دو رابطه‌ی بالا چه نتیجه‌ای می‌گیرید؟

[4] Efron-Stein inequality

۸ نامساوی ماکسیمال[5]

فرض کنید $\mathcal{T}$ یک خانواده‌ی متناهی از اندیس‌ها باشد. متغیرهای تصادفی $(X_t)_{t\in\mathcal{T}}$ دارای این ویژگی هستند که برای هر $t \in \mathcal{T}$ و برای هر $\lambda \geq 0$ داریم $\log \mathbf{E}\left[e^{\lambda X_t}\right] \leq \psi(\lambda)$ که در آن ψ یک تابع محدب است و $\psi(0) = \psi'(0) = 0$.

۱. ثابت کنید

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \leq \psi^{*-1}(\log |T|),$$

که در آن $\psi^*(x) = \sup_{\lambda\geq 0}\{\lambda x - \psi(\lambda)\}$ دوگان لژاندر[6] تابع ψ می‌باشد. حالت خاص این نامساوی در کلاس در حالی که X_tها σ^2-زیرگوسی باشند به صورت زیر ثابت شده است.

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \leq \sqrt{2\sigma^2 \log |T|}$$

۲. ثابت کنید

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq \psi^{*-1}(\log |T| + u)\right] \leq e^{-u} \quad \forall u \geq 0$$

حالت خاص این نامساوی در کلاس در حالی که X_tها σ^2-زیرگوسی باشند به صورت زیر ثابت شده است.

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq \sqrt{2\sigma^2 \log |T|} + x\right] \leq e^{-x^2/2\sigma^2} \quad \forall x \geq 0$$

۹ آنتروپی تجربی!

فرض کنید متغیر تصادفی X از یک توزیع احتمال گسسته و ناشناخته‌ی $P = (p_1, p_2, \ldots, p_k)$ می‌آید که $p_i = \mathbb{P}[X = a_i]$. هدف ما این است که تابع آنتروپی از این توزیع احتمال یعنی $H(P) = \sum_{i=1}^k p_i \log\left(\frac{1}{p_i}\right)$ را تخمین بزنیم. به جای دانستن توزیع احتمال P ما نمونه‌های مستقل $X_1, \ldots, X_n$ را از این توزیع مشاهده کرده‌ایم و هدف ما ساختن تخمین‌گری مانند $\hat{H}(X_1, \ldots, X_n)$ از روی نمونه‌ها برای $H(P)$ است.
یک راه برای ساختن این تخمین‌گر این است که از روی نمونه‌ها، توزیع را به صورت تجربی تخمین بزنیم و سپس از مقادیر تخمین‌زده شده برای هر احتمال استفاده کنیم. در حقیقت توزیع تجربی را به صورت $\hat{P}_n = (\hat{p}_1, \ldots, \hat{p}_k)$ تعریف می‌کنیم که در آن $\hat{p}_i = \frac{1}{n}\sum_{j=1}^n \mathbf{1}\{X_j = a_i\}$ و سپس تخمین‌گر خود را به صورت $\hat{H}(X_1, \ldots, X_n) = \hat{H}(\hat{P}_n) = \sum_{i=1}^k \hat{p}_i \log\left(\frac{1}{\hat{p}_i}\right)$ می‌سازیم.

۱. در این قسمت نشان خواهید داد $\hat{H}(\hat{P}_n)$ تخمین‌گری اریب است اما اریبی آن زیاد نیست. ثابت کنید

$$0 \leq H(P) - \mathbf{E}\left[\hat{H}(\hat{P}_n)\right] \leq \frac{k}{n}$$

۲. ثابت کنید $\hat{H}(X_1, \ldots, X_n) = \hat{H}(\hat{P}_n)$ خاصیت تفاضل محدود با $c_i = \frac{\log n}{n}$ دارد و درنتیجه متغیری $\frac{(\log n)^2}{n}$-زیرگوسی است.

۳. از دو قسمت قبل نتیجه بگیرید $\hat{H} \approx \mathbf{E}[\hat{H}] + \mathcal{O}(\frac{\log n}{\sqrt{n}})$ و $\mathbf{E}[\hat{H}] \approx H + \mathcal{O}(\frac{1}{n})$. از این دو چه نتیجه‌ای می‌گیرید؟

[5] maximal inequality
[6] Legendre dual

۱۰ توابع لیپشیتز (امتیازی)

در این مسئله قصد داریم نشان دهیم که توابع لیپشیتز از متغیرهای تصادفی گوسی، زیر گوسی هستند. تابع $f : \mathbb{R}^n \to \mathbb{R}$ را L- لیپشیتز نسبت به نرم اقلیدسی می‌گوییم اگر برای هر دو عضوی دلخواه x و y از دامنه داشته باشیم:

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

این گزاره برای توابع مشتق‌پذیر معادل است با $||\nabla f|| \leq L$. فرض کنید دو بردار X و Y دو بردار نرمال استاندارد مستقل با توزیع $\mathcal{N}(0, \mathbf{I}_{n\times n})$ باشند. می‌دانیم توزیع گوسی نسبت به دوران ناوردا است، پس به ازای هر $k \in [1 : n]$ تعریف می‌کنیم

$$Z_k(\theta) = X_k \sin(\theta) + Y_k \cos(\theta).$$

نکته: اگر تابع $g : \mathbb{R} \to \mathbb{R}$ لیپشیتز باشد، آن‌گاه تقریباً همه‌جا مشتق‌پذیر است.

۱. نشان دهید $Z(\theta)$ و مشتق آن نسبت به θ دو بردار گوسی و مستقل هستند.

۲. با توجه به این که $Z_k(0) = Y_k$ و $Z_k(\frac{\pi}{2}) = X_k$ است، می‌توان نوشت

$$f(X) - f(Y) = \int_0^{\frac{\pi}{2}} \frac{d}{d\theta} f\big(Z(\theta)\big) d\theta.$$

نشان دهید که برای هر تابع محدب $\phi : \mathbb{R} \to \mathbb{R}$ داریم

$$\mathbb{E}\big[\phi(f(X)) - \mathbb{E}[f(X)]\big] \leq \mathbb{E}\Big[\phi\big(\frac{\pi}{2}\langle \nabla f(X), Y\rangle\big)\Big]. \qquad (۱)$$

ج. قرار دهید $\phi(x) = e^{\lambda x}$ و نشان دهید که $f(X) - \mathbb{E}[f(X)]$ یک متغیر زیرگوسی با پارامتر حداکثر $\frac{\pi L}{2}$ می‌باشد.

د. به نظرتان آیا این نامساوی برای متغیرهای زیرگوسی هم برقرار است؟ سعی کنید کلّیت روش اثبات را ارائه داده یا یک مثال نقض بیاورید.

۱۱ مسئله واگذاری تکلیف (امتیازی)

مسئله واگذاری تکلیف (Assignment Problem) مسئله‌ای بسیار معروف است که واریانس جواب حالت تصادفی آن را بررسی می‌کنیم. فرض کنید یک ماتریس $X_{m\times m}$ داریم که درایه‌های آن متغیرهای تصادفی مستقل به طور یکنواخت روی بازه‌ی $[0, 1]$ هستند. مسئله بهنیه‌سازی زیر را درنظر بگیرید:

$$Z_m = \min_{\pi} \sum_{i=1}^{m} X_{i,\pi(i)}$$

که در آن مینیمم روی تمام جایگشتهای π روی مجموعه $\{1, \ldots, m\}$ است. سعی کنید با استفاده از کران واریانس و همچنین در نظر گرفتن گراف مجاورت برای ماتریس جایگشت π، نشان دهید

$$\mathrm{Var}(Z_m) = O(\frac{\log m}{m}).$$