به نام خدا

آنالیز احتمالاتی در ابعاد بالا (۱-۲۵۸۴۱)

تمرین سری دوم

ترم پاییز ۰۴-۱۴۰۳

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: جمعه ۹ آذر ساعت ۲۳:۵۹

(*) مسائلی که با ستاره مشخّص شده‌اند امتیازی هستند و حل‌کردن آن‌ها نمره‌ی امتیازی خواهد داشت!

۱ عدد پوششی

سه بخش این سؤال ارتباطی با یکدیگر ندارند.

۱. فرض کنید مجموعه‌ی $\mathcal{A}$ و متر d روی اعضای $\mathcal{A}$ داده شده باشد. مجموعه‌ی $\mathcal{K}$ را یک زیرمجموعه از $\mathcal{A}$ 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 $\mathcal{K}$، یعنی نقاط x_i، همگی اعضای مجموعه‌ی $\mathcal{K}$ باشند. عدد پوششی به‌دست‌آمده از این تعریف را $\mathcal{N}(\mathcal{K}, d, \epsilon)$ می‌نامیم که ϵ شعاع گوی‌های پوشش‌دهنده است.
- در تعریف دوم، مراکز گوی‌های پوشش دهنده، الزامی به عضو $\mathcal{K}$ بودن ندارند و می‌توانند هر عضو دلخواه $\mathcal{A}$ باشند. عدد پوششی به‌دست‌آمده از این تعریف را $\mathcal{N}^{\text{ext}}(\mathcal{K}, d, \epsilon)$ می‌نامیم.

به توجه به این دو تعریف، نشان دهید:

$$\mathcal{N}^{\text{ext}}(\mathcal{K}, d, \epsilon) \leq \mathcal{N}(\mathcal{K}, d, \epsilon) \leq \mathcal{N}^{\text{ext}}(\mathcal{K}, d, \frac{\epsilon}{۲}).$$

۲. در ابتدا، مثال نقضی برای قضیه‌ی شرطی زیر بزنید:

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

سپس تلاش نمایید قضیه‌ی شرطی زیر را ثابت کنید:

$$\mathcal{L} \subset \mathcal{K} \Longrightarrow \mathcal{N}(\mathcal{L}, d, \epsilon) \leq \mathcal{N}(\mathcal{K}, d, \frac{\epsilon}{۲}).$$

۳. مجموعه‌ی $\mathcal{K} = \{۰, ۱\}^n$ را با فاصله‌ی همینگ در نظر بگیرید. نشان دهید که برای هر عدد حسابی $m \leq n$ داریم:

$$\frac{۲^n}{\sum_{k=۱}^{m} \binom{n}{k}} \leq \mathcal{N}(\mathcal{K}, d_{\mathsf{H}}, m) \leq \mathcal{M}(\mathcal{K}, d_{\mathsf{H}}, m) \leq \frac{۲^n}{\sum_{k=۰}^{\lceil \frac{m}{۲} \rceil} \binom{n}{k}},$$

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

۴. مجموعه‌ی $\mathcal{K} = \{\circ, ۱\}^d$ و متر همینگ بهنجارشده‌ی $d_{\mathsf{H}}(\mathbf{x}, \mathbf{y}) = \frac{۱}{d}\sum_{j=۱}^{d} \mathbb{1}(x_j \neq y_j)$ را درنظر بگیرید. ثابت کنید عدد گنجایشی این فضای متری دارای کران بالای زیر است:

$$\frac{\log(\mathcal{M}(\mathcal{K}, d_{\mathsf{H}}, \delta))}{d} \leq D_{\mathrm{KL}}(\frac{\delta}{۲}\|۱/۲) + \frac{\log(d+۱)}{d}$$

که $D_{\mathrm{KL}}(\frac{\delta}{۲}||۱/۲) = \frac{\delta}{۲}\log(\frac{\delta/۲}{۱/۲}) + (۱ - \frac{\delta}{۲})\log(\frac{۱-\delta/۲}{۱/۲})$ همان آنتروپی نسبی باینری است.

۲ کران پایین برای ماکزیمم متغیرهای تصادفی

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

۱. اگر $A_۱, \ldots, A_n$ پیشامدهای مستقلی باشند، نشان دهید:

$$(۱ - e^{-۱})\Big\{۱ \wedge \sum_{k=۱}^{n} \mathbb{P}[A_k]\Big\} \leq \mathbb{P}\Big[\bigcup_{k=۱}^{n} A_k\Big] \leq ۱ \wedge \sum_{k=۱}^{n} \mathbb{P}[A_k].$$

که منظور از $a \wedge b$ همان $\min\{a, b\}$ است.

راهنمایی: دقت کنید که $\prod_{k=۱}^{n}(۱ - x_k) \leq \exp(-\sum_{k=۱}^{n} x_k)$ و $۱ - e^{-x} \geq (۱ - e^{-۱})(۱ \wedge x)$.

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

$$\mathbb{P}[X_t \geq x] \geq e^{-\eta^*(x)} \quad \forall x \geq \circ,\ \forall t \in \mathcal{T}$$

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

$$\mathbb{P}\Big[\sup_{t\in\mathcal{T}} X_t \geq \eta^{*-۱}(\log|T| + u)\Big] \geq (۱ - e^{-۱})e^{-u}.$$

حال که ما یک کران پایین روی احتمال دم ماکزیمم تعدادی متغیّر 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 $x \geq \circ$ داریم:

$$\mathbb{P}\Big[\sup_{t\in\mathcal{T}} X_t \geq \frac{۱}{۲}\eta^{*-۱}(۲\log|T|) + x)\Big] \geq (۱ - e^{-۱})\exp(\frac{-\eta^*(۲x)}{۲}).$$

راهنمایی: از مقعّر بودن $\eta^{*-۱}$ استفاده کنید.

۴. فرض کنید ψ^* نیز یک تابع اکیدً صعودی و محدّب باشد. هم‌چنین فرض کنید که:

$$e^{-\eta^*(x)} \leq \mathbb{P}[X_t \geq x] \leq e^{-\psi^*(x)} \quad \forall x \geq \circ,\ \forall t \in \mathcal{T}.$$

حال نتیجه بگیرید که ثابت‌های مثبت $C_۱$ و $C_۲$ وجود دارند، به نحوی که:

$$C_۱\left(\eta^{*-۱}(\log|\mathcal{T}|) + \sup_{t\in\mathcal{T}} \mathbb{E}[\circ \wedge X_t]\right) \leq \mathbb{E}\Big[\sup_{t\in\mathcal{T}} X_t\Big] \leq C_۲(\psi^{*-۱}(\log|\mathcal{T}|))$$

راهنمایی: از $\mathbb{E}[\max\{\circ, Z\}] = \int_{\circ}^{\infty} \mathbb{P}[Z \geq x]dx$ استفاده کنید.

کران‌های بالا و پایین که در این قسمت به دست آورده‌ایم، معمولاً از یک مرتبه هستند به شرط اینکه کران‌های بالا و پایینی که در شروع روی $\mathbb{P}[X_t \geq x]$ می‌گذاریم از یک مرتبه باشند. به عنوان مثال متغیّرهای تصادفی گوسی را بررسی می‌کنیم.

۵. برای متغیر تصادفی $X \sim \mathcal{N}(۰,۱)$ نشان دهید:

$$\mathbb{P}[X \geq x] \geq \frac{e^{-x^۲}}{۲\sqrt{۲}} \quad \forall x \geq ۰.$$

راهنمایی: احتمال را به صورت انتگرالی نوشته و از نامساوی $(x+v)^۲ \leq ۲v^۲ + ۲x^۲$ بهره ببرید.

۶. فرض کنید $X_۱, \ldots, X_n$ متغیّرهای تصادفی مستقل و هم‌توزیع با توزیع نرمال استاندارد باشند. از قسمت‌های قبل کمک گرفته و نشان دهید:

$$\frac{۱-e^{-۱}}{۲}\sqrt{۲\log(۲^{-۳/۴}n)} - \frac{۱}{\sqrt{۲\pi}} \leq \mathbb{E}\Big[\max_{i\leq n} X_i\Big] \leq \sqrt{۲\log(n)}$$

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

۳ مسئله‌ی فروشنده‌ی دوره‌گرد

فرض کنید $\mathbf{X}_۱, \ldots, \mathbf{X}_n$ نقاطی تصادفی و مستقل و هم توزیع باشند که به صورت یکنواخت روی مربّع واحد $[۰,۱]^۲$ توزیع شده‌اند. ما به $\mathbf{X}_i$ به چشم مکان شهر iام در مربّع واحد نگاه می‌کنیم. هدف مسئله‌ی فروشنده‌ی دوره‌گرد پیدا کردن گشتی است که یک بار از همه‌ی شهرها بگذرد و کمترین مسافت ممکن را طی کرده باشد. ما طول کمترین گشت را به صورت:

$$L_n := \min_\sigma \left\{\|\mathbf{X}_{\sigma(۱)} - \mathbf{X}_{\sigma(۲)}\| + \|\mathbf{X}_{\sigma(۲)} - \mathbf{X}_{\sigma(۳)}\| + \cdots + \|\mathbf{X}_{\sigma(n)} - \mathbf{X}_{\sigma(۱)}\|\right\},$$

نشان می‌دهیم، که در آن کمینه‌سازی روی همه‌ی جایگشت‌های ممکن از $\{۱, \ldots, n\}$ است. برای شروع به میزان بزرگی L_n می‌پردازیم.

۱. نشان دهید $\mathbb{E}[L_n] \asymp \sqrt{n}$.

راهنمایی: برای کران پایین ثابت کنید که: $L_n \geq \sum_{k=۱}^{n} \min_{l\neq k} \|\mathbf{X}_k - \mathbf{X}_l\|$ و برای کران بالا ثابت کنید که: $L_n \leq L_{n-۱} + ۲\min_{k<n}\|\mathbf{X}_n - \mathbf{X}_k\|$.

۲. از نامساوی McDiarmid استفاده کنید و نشان دهید L_n یک متغیّر تصادفی $۲n$-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 حتّی نتیجه نمی‌دهد که L_n حول متوسّطاش 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.

نامساوی Talagrand: فرض کنید $X_۱, \ldots, X_n$ متغیّرهای تصادفی مستقل باشند و داشته باشیم:

$$f(\mathbf{x}) - f(\mathbf{y}) \leq \sum_{i=۱}^{n} c_i(x)\mathbb{1}(x_i \neq y_i) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

در این صورت $f(X_۱, \ldots, X_n)$ یک متغیّر تصادفی $\|\sum_{i=۱}^{n} c_i^۲\|_\infty$-زیرگوسی است.

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

۳. فرض کنید $\mathbf{v} = (۰, a)$ و $\mathbf{w} = (b, ۰)$ رئوس مثلّث قائم‌الزاویه‌ی $\mathcal{T} = \mathrm{conv}\{۰, \mathbf{v}, \mathbf{w}\}$ باشند. نشان دهید

$$\forall \mathbf{x} \in \mathcal{T}: \quad \|\mathbf{v} - \mathbf{x}\|^۲ + \|\mathbf{x} - \mathbf{w}\|^۲ \leq \|\mathbf{v} - \mathbf{w}\|^۲.$$

۴. اثبات کنید که برای هر $\mathbf{x}_۱, \ldots, \mathbf{x}_n \in \mathcal{T}$ یک جایگشت مانند σ از $\{۱, ۲, \ldots, n\}$ وجود دارد به نحوی که

$$\|\mathbf{v} - \mathbf{x}_{\sigma(۱)}\|^۲ + \sum_{i=۱}^{n-۱} \|\mathbf{x}_{\sigma(i)} - \mathbf{x}_{\sigma(i+۱)}\|^۲ + \|\mathbf{x}_{\sigma(n)} - \mathbf{w}\|^۲ \leq \|\mathbf{v} - \mathbf{w}\|^۲.$$

راهنمایی: از استقرا استفاده کنید. فرض کنید حکم برای هر مثلّث قائم‌الزاویه مانند $\mathcal{S}$ و $\mathbf{x}_1, \dots, \mathbf{x}_{n-1} \in \mathcal{S}$ برقرار است. مثلّث $\mathcal{T}$ را با رسم ارتفاع از مبدأ به وتر به دو مثلّث قائم‌الزاویه‌ی کوچک‌تر تقسیم کنید. اگر هر کدام از زیرمثلّث‌ها شامل نقاطی بودند از فرض استقرا استفاده کنید تا حکم را نتیجه بگیرید. در غیر این صورت آن‌قدر به تقسیم‌کردن مثلّث‌ها ادامه دهید تا بتوان از فرض استقرا استفاده کرد.

۵. حال نتیجه بگیرید که برای هر مجموعه نقاط مانند $\mathbf{x}_1, \dots, \mathbf{x}_n \in [0,1]^2$ یک جایگشت مانند σ از $\{1, 2, \dots, n\}$ وجود دارد به نحوی که:

$$\|\mathbf{x}_{\sigma(1)} - \mathbf{x}_{\sigma(2)}\|^2 + \|\mathbf{x}_{\sigma(2)} - \mathbf{x}_{\sigma(3)}\|^2 + \cdots + \|\mathbf{x}_{\sigma(n)} - \mathbf{x}_{\sigma(1)}\|^2 \leq 4.$$

حالا می‌توانیم از یافته‌های هندسی خود برای تحلیل طول گشت فروشنده‌ی دوره‌گرد استفاده کنیم. به یاد داریم که یک گشت بین نقاط $\mathbf{x}_1, \dots, \mathbf{x}_n$ به کمک یک جایگشت σ از $\{1, 2, \dots, n\}$ تعریف می‌شد. طول یک گشت مشخص را با نماد $l_n(\{\mathbf{x}_i\}_{i=1}^n, \sigma)$ نشان می‌دهیم. در نتیجه: $L_n := \min_\sigma l_n(\{X_i\}_{i=1}^n, \sigma)$.

۶. فرض کنید $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ و $\mathcal{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ دو مجموعه نقاطی در $[0,1]^2$ باشند که $\mathcal{X} \cap \mathcal{Y} \neq \emptyset$. فرض کنید σ یک گشت برای $\mathcal{X}$ و τ یک گشت برای $\mathcal{Y}$ باشد. نشان دهید یک گشت مانند ρ از $\mathcal{X} \cup \mathcal{Y}$ وجود دارد به نحوی که:

$$l_{2n}(\mathcal{X} \cup \mathcal{Y}, \rho) \leq l_n(\mathcal{Y}, \tau) + 2\sum_{i=1}^{n} \mathbb{1}(\mathbf{x}_i \notin \mathcal{Y}) d_i(\mathcal{X}, \sigma),$$

که در آن $d_i(\mathcal{X}, \sigma)$ فاصله‌ی بین $\mathbf{x}_i$ و نقطه‌ی قبلی در گشت σ است.
راهنمایی: فرض کنید σ و τ دو مسیر پیاده‌روی هستند که با رنگ‌های قرمز و آبی مشخص شده‌اند و در بعضی شهرها با هم تلاقی دارند. هدف شما این است که به طور سیستماتیک اجتماع مسیرها را پیاده‌روی کنید. برای این منظور، پیاده‌روی 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 پیاده‌روی 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 $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset [0,1]^2$ گشت $\sigma_{\mathcal{X}}$ را همان گشتی در نظر بگیرید که خاصیت قسمت ۵ را دارد. نشان دهید برای هر $\mathcal{X}, \mathcal{Y} \subset [0,1]^2$ که $|\mathcal{X}| = |\mathcal{Y}| = n$ داریم:

$$\min_\sigma l_n(\mathcal{X}, \sigma) \leq \min_\sigma l_n(\mathcal{Y}, \sigma) + \sum_{i=1}^{n} 2 d_i(\mathcal{X}, \sigma_{\mathcal{X}}) \mathbb{1}(\mathbf{x}_i \neq \mathbf{y}_i).$$

۸. تیجه بگیرید که L_n برای هر $n \geq 1$ یک متغیر ۱۶−−زیرگوسی است.

۴ نرم ماتریس زیرگوسی

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

$$\|\frac{1}{m}\mathbf{A}^\top \mathbf{A} - \mathbf{I}_n\|_{\mathsf{op}} \leq C \max\{\delta, \delta^2\},$$

که در آن $\delta = \sqrt{\frac{n}{m}} + \frac{t}{\sqrt{m}}$ است.
یادداشت: بردار تصادفی $\mathbf{X}$ زیرگوسی است اگر به ازای هر بردار $\mathbf{v}$ عضو کره‌ی واحد، $Z = \mathbf{v}^\top \mathbf{X}$ یک متغیّر تصادفی زیرگوسی با ثابت محدود باشد. در قضیه فوق، ثابت C می‌توانند به ضرایب زیرگوسی بودن بردار $\mathbf{X}$ نیز وابسته باشند.
راهنمایی: نرم اپراتوری را به فرم سوپریمی آن نوشته و تلاش کنید با انتخاب یک $\epsilon - \mathrm{net}$ 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 chaining نیست.

۵ کران روی واریانس ماکسیمم گوسی‌ها

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

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

راهنمایی: بردار تصادفی $\mathbf{X}$ را به صورت $\mathbf{X} = \mathbf{C}^{\frac{1}{2}}\mathbf{Y}$ بنویسید که $Y_1, \ldots, Y_n$ متغیّرهای تصادفی گوسی استاندارد و مستقل از هم هستند.

۶ $(*)$ بسته‌بندی بارهای کشتی!

به عنوان راهنمایی در شروع می‌گوییم که این سوال یکی از کاربردهای قدیمی و معروف نامساوی افرون_اشتاین است.

۱. برای شروع نتیجه‌ی مهم زیر از نامساوی افرون_اشتاین را ثابت کنید:

$$\mathrm{Var}[f(X_1, \ldots, X_n)] \leq \frac{1}{4}\,\mathbb{E}\Big[\sum_{i=1}^{n} \big(D_i f(X_1, \ldots, X_n)\big)^2\Big],$$

که در آن:

$$D_i f(x_1, \ldots, x_n) := \sup_z f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) - \inf_z f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n).$$

۲. حال فرض کنید که $X_1, \ldots, X_n$ متغیّرهای تصادفی مستقل و هم توزیع، با مقادیر در بازه‌ی $[\circ, 1]$، و با امید ریاضی $\frac{1}{2}$ باشند. هر کدام از این متغیّرهای تصادفی نشان‌دهنده‌ی اندازه‌ی میزان باری است که باید توسّط یک کشتی جابه‌جا شود. هر کدام از این بارها باید در محفظه‌هایی به اندازه‌ی ۱ جاسازی شوند تا بتوانند داخل کشتی انبار شده و سپس حمل شوند. بنابراین هر محفظه می‌تواند مجموعه‌ای از بارها که جمع اندازه‌شان حدّاکثر ۱ است را در داخل خود جای دهد. $B_n = f(X_1, \ldots, X_n)$ را کمترین تعداد محفظه‌هایی تعریف می‌کنیم که بارها همگی در آنها جاسازی شوند. جالب است بدانید محاسبه‌ی B_n یک مسئله‌ی بهینه‌سازی ترکیبیاتی سخت به حساب می‌آید ولی می‌توان میانگین و واریانس آن را به سادگی کران زد.

- نشان دهید: $\mathrm{Var}[B_n] \leq \frac{n}{4}$.
- نشان دهید: $\mathbb{E}[B_n] \geq \frac{n}{2}$.

از دو رابطه‌ی بالا چه نتیجه‌ای می‌گیرید؟

۷ $(*)$ نامساوی ماکسیمال

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

۱. ثابت کنید:

$$\mathbb{E}\Big[\sup_{t\in\mathcal{T}} X_t\Big] \leq \psi^{*-1}(\log|\mathcal{T}|),$$

که در آن $\psi^*(x) = \sup_{\lambda\geq\circ}\left\{\lambda x - \psi(\lambda)\right\}$ دوگان لژاندر تابع ψ است.

۲. به ازای هر $u \geq \circ$ ثابت کنید:

$$\mathbb{P}\Big[\sup_{t\in\mathcal{T}} X_t \geq \psi^{*-1}(\log|\mathcal{T}| + u)\Big] \leq e^{-u}.$$

۸ کران بالا برای پیچیدگی گوسی

مجموعه‌ی $\mathcal{T}(s) = \{\theta \in \mathbb{R}^d \mid \|\theta\|_۰ \leq s, \|\theta\|_۲ \leq ۱\}$ را در نظر بگیرید. این مجموعه شامل تمام بردارهای s–تُنُک است که در داخل یک توپ با شعاع واحد قرار دارد. در این سوال می‌خواهیم یک کران بالا بر روی پیچیدگی گوسی این مجموعه، $\mathcal{G}(\mathcal{T}(s))$ به دست آوریم. برای یک مجموعه مانند $\mathcal{T} \subset \mathbb{R}^d$ پیچیدگی گوسی به صورت زیر تعریف می‌شود:

$$\mathcal{G}(\mathcal{T}) = \mathbb{E}\left[\sup_{\mathbf{x}\in\mathcal{T}} \left|\mathbf{x}^\top \mathbf{Z}\right|\right],$$

که $\mathbf{Z} \sim \mathcal{N}(۰, \mathbf{I}_d)$.

۱. بردار $\mathbf{w}_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}$ را در نظر بگیرید که یک زیربردار از $(w_۱, \ldots, w_d)$ است و اندیس‌های آن از $\mathcal{S} \subset \{۱, ۲, \ldots, d\}$ می‌آید. نشان دهید:

$$\mathcal{G}(\mathcal{T}(s)) = \mathbb{E}\Big[\max_{|\mathcal{S}|=s} \|\mathbf{w}_\mathcal{S}\|_۲\Big].$$

۲. برای هر زیر مجموعه‌ی $\mathcal{S}$ با اندازه‌ی s نشان دهید:

$$\mathbb{P}[\|\mathbf{w}_\mathcal{S}\|_۲ \geq \sqrt{s} + \delta] \leq e^{\frac{-\delta^۲}{۲}}.$$

۳. با استفاده از قسمت‌های قبل نشان دهید:

$$\mathcal{G}(\mathcal{T}(s)) \leq c\sqrt{s \log(\frac{ed}{s})}.$$

۹ $(*)$ توابع لیپشیتز

در این مسئله قصد داریم نشان دهیم که توابع لیپشتیز از متغیّرهای تصادفی گوسی، زیر گوسی هستند. تابع $f : \mathbb{R}^n \mapsto \mathbb{R}$ را L–لیپشیتز نسبت به نرم اقلیدسی می‌گوییم اگر برای هر دو عضو دلخواه $\mathbf{x}, \mathbf{y}$ از دامنه داشته باشیم:

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

این گزاره برای توابع مشتق‌پذیر معادل است با: $\|\nabla f\| \leq L$. فرض کنید دو بردار $\mathbf{X}, \mathbf{Y}$ دو بردار نرمال استاندارد مستقل با توزیع $\mathcal{N}(۰, \mathbf{I}_n)$ باشند. می‌دانیم توزیع گوسی نسبت به دوران ناوردا است، به ازای هر $k \in [۱ : n]$ تعریف می‌کنیم:

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

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

۱. نشان دهید $Z_k(\theta)$ و مشتق آن نسبت به θ دو متغیّر تصادفی گوسی و مستقل هستند.

۲. با توجه به اینکه $Z_k(۰) = Y_k$ و $Z_k(\frac{\pi}{۲}) = X_k$ است، می‌توان نوشت:

$$f(\mathbf{X}) - f(\mathbf{Y}) = \int_۰^{\frac{\pi}{۲}} \frac{d}{d\theta} f\big(\mathbf{Z}(\theta)\big) d\theta.$$

نشان دهید که برای هر تابع محدّب $\phi : \mathbb{R} \mapsto \mathbb{R}$ داریم:

$$\mathbb{E}\left[\phi\big(f(\mathbf{X})\big) - \mathbb{E}[f(\mathbf{X})]\right] \leq \mathbb{E}\left[\phi\big(\frac{\pi}{۲}\mathbf{Y}^\top \nabla f(\mathbf{X})\big)\right].$$

۳. قرار دهید $\phi(x) = e^{\lambda x}$ و نشان دهید که $f(\mathbf{X}) - \mathbb{E}[f(\mathbf{X})]$ یک متغیّر تصادفی زیرگوسی با پارامتر حدّاکثر $\frac{\pi L}{۲}$ است.

۴. به نظرتان آیا این نامساوی برای متغیّرهای تصادفی زیرگوسی هم برقرار است؟ سعی کنید کلیت روش اثبات را ارائه داده یا یک مثال نقض بیاورید.