به نام خدا

نظریه‌ی اطّلاعات، آمار و یادگیری (۱-۲۵۱۱۰)

تمرین سری چهارم

ترم بهار ۰۳-۱۴۰۲

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: جمعه ۲۵ خرداد ۱۴۰۳ ساعت ۲۳:۵۹

۱ متغیّر یکنواخت، متغیّر مهربان!

متغیّر تصادفی یکنواخت U_θ با پارامتر θ، متغیّری یکنواخت روی بازه‌ی $[\theta, \theta + ۱]$ است. هدف تخمین زدن پارامتر θ تحت تابع زیان مجذور فاصله از روی n مشاهده‌ی i.i.d. از U_θ است.

۱. با استفاده از روش دونقطه‌ای Le-Cam نشان دهید:

$$\inf_{\hat\theta} \sup_{\theta\in\mathbb{R}} \mathbb{E}[(\theta - \hat\theta)^۲] \geq \frac{c}{n^۲}.$$

۲. یک تخمین‌گر برای این مسئله ارائه کنید که به کران فوق می‌رسد. یعنی تخمین‌گر $\hat\theta$ را طوری تعیین کنید که:

$$\sup_{\theta\in\mathbb{R}} \mathbb{E}[(\theta - \hat\theta)^۲] \leq \frac{C}{n^۲}.$$

۳. آیا کران فوق با کران‌های معمول برای مسائل مشابه هم‌خوانی دارد؟ توضیح دهید.

۲ f-انحراف و نامساوی فانو

هدف از این مسئله ارائه‌ی تعمیمی از نامساوی فانو است. فرض کنید M متغیّری تصادفی با توزیع یکنواخت روی $[۱ : m]$ باشد (مثلا در مسئله‌ی آزمون فرض بیزی با توزیع پیشین یکنواخت). X مشاهده‌ی ما از فرض ناپیدای M است و توزیع مشترک P_{MX} داده شده است. ضمنا یک قاعده‌ی تصمیم‌گیری تصادفی $T_{\hat{M}|X}$ داده شده است و فرض کنید که احتمال خطا برابر با P_E باشد.

۱. Q_X را توزیع احتمال دلخواهی در نظر بگیرید. نشان دهید:

$$D_f\left(P_{MX}T_{\hat{M}|X} \middle\| P_M Q_X T_{\hat{M}|X}\right) = D_f\left(P_{MX} \| P_M Q_X\right).$$

۲. نشان دهید:

$$D_f\left(P_{MX} \| P_M Q_X\right) \geq D_f\left(P_{M\hat{M}} \| P_M Q_{\hat{M}}\right)$$

که در آن $P_{M\hat{M}}$ توزیع حاشیه‌ای القا شده از توزیع P_{MX} و قاعده‌ی تصمیم‌گیری است.

۳. نشان دهید:

$$D_f\left(P_{M\hat{M}} \| P_M Q_{\hat{M}}\right) \geq D_f\left(\mathsf{Ber}(P_E) \,\|\, \mathsf{Ber}(۱ - \frac{۱}{m})\right)$$

۴. در نهایت نتیجه بگیرید:

$$\inf_{Q_X} \max_m D_f\left(P_{X|M=m} \| Q_X\right) \geq \inf_{Q_X} D_f\left(P_{MX} \| P_M Q_X\right) \geq D_f\left(\mathsf{Ber}(P_E) \,\|\, \mathsf{Ber}(۱ - \frac{۱}{m})\right)$$

۳ نقطه در برابر مخلوط!

در روش Le-Cam دو نقطه از فضا به دیدار هم می‌رفتند و مسئله به یک آزمون فرض بیزی تبدیل می‌شد. در این مسئله تعمیم ساده‌ای از این روش را بیان می‌کنیم که در آن یک نقطه به دیدار مجموعه‌ای از نقاط دیگر می‌رود.

۱. به صورت دقیق‌تر، فرض کنید برای تابع زیان $\ell : \Theta \times \hat{\Theta} \to \mathbb{R}$، نقطه‌ی $\theta_۰$ و مجموعه‌ی $\Theta_۱ \subseteq \Theta$ موجود باشند به طوری‌که شرط زیر برقرار باشد:

$$\ell(\theta_۰, a) + \ell(\theta_۱, a) \geq \Delta, \quad \forall a \in \hat{\Theta},\ \theta_۱ \in \Theta_۱.$$

نشان دهید برای هر توزیع احتمال دلخواه μ روی $\Theta_۱$، کران زیر برقرار است:

$$\inf_{\hat{\theta}} \sup_{\theta \in \Theta} \ \mathbb{E}_\theta\left[\ell(\theta, \hat{\theta})\right] \geq \frac{\Delta}{۲}\left(۱ - d_{\mathrm{TV}}\left(P_{\theta_۰}, \mathbb{E}_{\mu_{\theta_۱}} P_{\theta_۱}\right)\right).$$

دقت کنید که $\mathbb{E}_{\mu_{\theta_۱}} P_{\theta_۱}$ یک توزیع احتمال مخلوط است.

۲. در حالت کلی محاسبه کران بالا به خصوص محاسبه d_{TV} بین توزیع مخلوط ضریبی و یک توزیع دیگر چالش برانگیز است. به همین منظور، از کرانی مبتنی بر فاصله‌ی $\chi^۲$ استفاده می‌شود. ابتدا نشان دهید:

$$d_{\mathrm{TV}}(P, Q) \leq c\ D_{\chi^۲}(P, Q).$$

مقدار c چقدر است؟

۳. نشان دهید:

$$۱ + D_{\chi^۲}(\mathbb{E}_\mu[P_\theta], Q) = \mathbb{E}\left[\int \frac{P_\theta(x) P_{\theta'}(x)}{Q(x)} dx\right].$$

که در آن θ, θ' مستقل از یکدیگر و از روی توزیع μ تولید شده‌اند.

۴. حال می‌خواهیم به کمک کران به دست آمده مسئله‌ی تخمین ماکزیمم متغیّرهای تصادفی گوسی را بررسی کنیم.

دنباله‌ی گوسی $X_i = \theta_i + Z_i$ را برای $i \in [n]$ در نظر بگیرید، که بردار پارامتر $\theta = (\theta_۱, \theta_۲, \ldots, \theta_p)$ می‌تواند هر مقداری از $\mathbb{R}^n$ بگیرد. همچنین $Z_۱, Z_۲, \ldots, Z_p \sim \mathcal{N}(۰, ۱)$ نمونه‌هایی به صورت i.i.d هستند. فرض کنید می‌خواهیم $\theta_{\max} = \max_{i \in [n]} \theta_i$ را با تخمین‌گر T، تحت تابع هزینه‌ی $L(\theta, T) = (T - \theta_{\max})^۲$ تخمین بزنیم. نشان دهید ثابت $c \geq ۰$ و مستقل از n موجود است، که:

$$\inf_T \sup_{\theta \in \mathbb{R}^n} \ \mathbb{E}_\theta[L(\theta, T)] \geq c \cdot \log n.$$

۵. در مسئله‌ی قسمت قبل، تخمین‌گر T را به گونه‌ای ارائه دهید که:

$$\sup_{\theta \in \mathbb{R}^n} \ \mathbb{E}_\theta[L(\theta, T)] \leq C \cdot \log n,$$

برای یک $C \leq \infty$ که مستقل از n است.

۴ Chernoff-Rubin-Stein

تعداد n مشاهده‌ی X_i به صورت i.i.d. از توزیع P_θ داریم و می‌دانیم $\theta \in [-a, a]$.

۱. نشان دهید:

$$R_n^*(\Theta) \triangleq \inf_{\hat{\theta}} \sup_{\theta \in [-a,a]} \mathbb{E}_\theta \left[\|\hat{\theta} - \theta\|_2^2 \right] \geq \min_{0<\epsilon<1} \max \left(\epsilon^2 a^2, \frac{1 - \epsilon^2}{n\overline{I}} \right).$$

که $\overline{I} = \frac{1}{2a} \int_{-a}^{a} I(\theta)\, d\theta$ میانگین اطّلاعات فیشر است.

۲. با ساده کردن نامساوی قسمت قبل نتیجه بگیرید:

$$R_n^*(\Theta) \geq \left(\frac{1}{a^{-1} + \sqrt{n\overline{I}}} \right)^2$$

۳. فرض پیوستگی $I(\theta)$ بر حسب θ را در نظر بگیرید و با این فرض نشان دهید که نابرابری قسمت قبل، رابطه‌ی زیر را نتیجه می‌دهد:

$$\inf_{\hat{\theta}} \sup_{\theta \in \left[\theta_0 - n^{-\frac{1}{4}}, \theta_0 + n^{-\frac{1}{4}}\right]} \mathbb{E}_\theta \left[\|\hat{\theta} - \theta\|_2^2 \right] \geq \frac{1 + O(1)}{nI(\theta_0)}$$

۵ تخمین توزیع

فرض کنید تعداد n نمونه‌ی $X_1, \ldots, X_n$ به شکل مستقل از توزیع P روی مجموعه‌ی $[k]$ انتخاب شده‌اند. هدف در این سوال این است که نشان دهیم خطای min–max برای تخمین P در شرط زیر صدق می‌کند:

$$\inf_{\hat{P}} \sup_{P \in \mathcal{M}_k} \mathbb{E}_P[d_{\mathrm{TV}}(\hat{P}, P)] \asymp \min \left\{ \sqrt{\frac{k-1}{n}}, 1 \right\}, \quad k, n \in \mathbb{N}.$$

۱. نشان دهید که تخمین MLE نرخ پیشنهادی بالا را دارد.

۲. یک کران پایین برای خطای min–max با استفاده از روش Assouad به دست آورید.

۳. یک کران پایین برای خطای min–max با استفاده از روش Fano به دست آورید.