به نام خدا

نظریه‌ی اطّلاعات، آمار و یادگیری (۱-۲۵۱۱۰)

تمرین سری دوم

ترم بهار ۰۳-۱۴۰۲

دانشکده‌ی مهندسی برق

دانشگاه صنعتی شریف

استاد: دکتر محمّدحسین یاسائی میبدی

مهلت تحویل: جمعه ۷ اردی‌بهشت ۱۴۰۳ ساعت ۲۳:۵۹

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

۱ انحراف بزرگ برای Log-Likelihood

فرض کنید P و Q دو توزیع احتمال باشند که $P \ll Q$. همین‌طور X_iها متغیّرهای تصادفی i.i.d از توزیع P و Y_iها متغیّرهای تصادفی i.i.d. از توزیع Q هستند. تعریف می‌کنیم: $Z_i = \log\left(\frac{p(X_i)}{q(X_i)}\right)$ و $W_i = \log\left(\frac{p(Y_i)}{q(Y_i)}\right)$.
در این سوال می‌خواهیم رابطه‌ی زیر را به ازای هر $t \geq ۰, n \in \mathbb{N}$ ثابت کنیم:

$$\mathbb{P}\left[\sum_{i=۱}^{n}(W_i - Z_i) \geq nt\right] \leq \exp\left(-n\left(\alpha + \frac{t}{۲}\right)\right)$$

که در آن $\alpha = -۲\log\mathcal{B}(P,Q)$ و $\mathcal{B}(P,Q) = \mathbb{E}_{Y\sim Q}\left[\sqrt{\frac{p(Y)}{q(Y)}}\right]$.

۱. ثابت کنید:

$$\mathbb{P}\left[\sum_{i=۱}^{n}(W_i - Z_i) \geq nt\right] \leq \exp\left(-n \cdot F(t)\right),$$

که در آن: $F(t) = \sup_{\lambda\geq ۰}\{\lambda t - \psi_P(-\lambda) - \psi_Q(\lambda)\}$، $\psi_P(\lambda) = \log\mathbb{E}\left[e^{\lambda Z_i}\right]$ و $\psi_Q(\lambda) = \log\mathbb{E}\left[e^{\lambda W_i}\right]$.

۲. ثابت کنید: $F(۰) = -\psi_P(-\frac{۱}{۲}) - \psi_Q(\frac{۱}{۲}) = \alpha$

۳. ثابت کنید: $F(t) \geq F(۰) + \frac{t}{۲}$ سپس حکم را نتیجه بگیرید.

۲ زوج نرخ‌های قابل دسترس

در درس، دیدیم یک روش بررسی رفتار حدی خطاهای مسئله‌ی آزمون فرض آن است که خطای $\pi_{۱|۰}$ را کوچک نگه داریم و نرخ‌های همگرایی قابل دسترس برای خطای $\pi_{۰|۱}$ را به دست آوریم. حال در این مسئله می‌خواهیم برای هر دو عبارت خطا نرخ همگرایی به دست آوریم. منحنی مرزی ناحیه‌ی زوج نرخ‌های همگرایی قابل دسترس یعنی زوج نرخ‌هایی مانند $E_۰$ و $E_۱$ که برای آن‌ها روش تصمیم‌گیری‌ای وجود دارد که در آن داریم:

$$\pi_{۱|۰} \leq ۲^{-n\cdot E_۰}, \pi_{۰|۱} \leq ۲^{-n\cdot E_۱}$$

۱. استدلال کنید که چرا ناحیه‌ی زوج نرخ‌های قابل دسترس باید یک ناحیه‌ی محدّب باشد؟

۲. با استفاده از قضیه‌ی Neyman-Pearson و قرار دادن $\tau = n \cdot t$، که τ پارامتر روش تصمیم گیری LLR است، نشان دهید به شرط $-D_{\mathrm{KL}}(P\|Q) \le t \le D_{\mathrm{KL}}(Q\|P)$ داریم:

$$\pi_{۱|۰}^{(n)} \le ۲^{-n\cdot\psi_P^*(t)}, \qquad \pi_{۰|۱}^{(n)} \le ۲^{-n\psi_Q^*(t)}.$$

که: $\psi_P^*(t) = \sup_{\lambda\in\mathbb{R}}\{\lambda t - \psi_P(\lambda)\}$ و $\psi_P(\lambda) = \log \mathbb{E}_{X\sim P}\left[\exp\left(\lambda \log \frac{p(X)}{q(X)}\right)\right]$.

۳. با استفاده از نامساوی‌های فوق نشان دهید که به ازای هر t که در شرط $-D_{\mathrm{KL}}(P\|Q) \le t \le D_{\mathrm{KL}}(Q\|P)$ صدق کند، زوج نرخ زیر قابل‌حصول هستند:

$$E_۰(t) = \psi_P^*(t), \qquad E_۱(t) = \psi_P^*(t) - t.$$

۴. حال نشان دهید که منحنی پارامتری $\begin{cases} E_۰(t) = \psi_P^*(t) \\ E_۱(t) = \psi_P^*(t) - t \end{cases}$، همان منحنیِ مرزیِ زوج‌های قابل‌حصول است.

۵. هدف آنست که نرخ بهینه‌ی همگرایی عبارت زیر را محاسبه کنیم:

$$\min_{P(Z|X^n)} \left\{\pi_۰\pi_{۱|۰} + \pi_۱\pi_{۰|۱}\right\} \tag{۱}$$

با استفاده از مرزی که برای ناحیه‌ی زوج نرخ‌های همگرایی قابل‌حصول به دست آوردیم، مسئله‌ی محاسبه‌ی نرخ بهینه‌ی همگرایی عبارت (۱) به ازای مقادیر ثابت احتمال‌های اوّلیه‌ی $\pi_۰, \pi_۱$ را به صورت یک مسئله‌ی max-min درآورید و نشان دهید نرخ بهینه برابر است با $\psi_P^*(۰)$.

۳ گریز از مرکز

گوی به شعاع r در فضای $\mathbb{R}^n$ را چنین تعریف می‌کنیم:

$$\mathcal{B}^{(n)}(r) = \{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\|_۲ \le r\}.$$

مجموعه‌ی $\mathcal{A}^{(n)}$ را چنین تعریف می‌کنیم:

$$\mathcal{A}^{(n)} = \{\mathbf{x} \in \mathbb{R}^n \ : \ \forall i \in [n]\ x_i \in \{۱, ۲\}\}.$$

به تعبیر دیگر، $\mathcal{A}^{(n)}$ مجموعه‌ی نقاطی از فضای $\mathbb{R}^n$ است که هرکدام از درایه‌های آن‌ها یکی از اعداد ۱ یا ۲ باشد. فرض کنید $r_۰$ یک عدد ثابت مثبت باشد. تعداد نقاط $\mathcal{A}^{(n)}$ که خارج از گوی به شعاع $r_۰\sqrt{n}$ باشند را با $f(r_۰, n)$ نشان می‌دهیم. نشان دهید عددی مثل r^* وجود دارد که اگر $r_۰ < r^*$ باشد، آن‌گاه $f(r_۰, n)$ به عنوان تابعی از n نمای برابر ۱ دارد و اگر $r_۰ > r^*$ باشد، آنگاه نمای تابع $f(r_۰, n)$ به عنوان تابعی از n اکیداً کمتر از ۱ است.
منظور از نمای یک تابع مانند $h(n)$ مقدار عبارت $\lim\limits_{n\to\infty} \frac{\log(h(n))}{n}$ است.

۴ اطّلاعات چرنف

فرض کنید نمونه‌های $X_۱, X_۲, \ldots, X_n$ به صورت .i.i.d از توزیع Q به ما داده شده باشد. حالت بیزی را در نظر بگیرید که می‌دانیم با احتمال اوّلیه‌ی $\pi_۱$، داریم $Q = P_۱$ و همچنین با احتمال اوّلیه‌ی $\pi_۲$ داریم $Q = P_۲$. همچنین فرض کنید $\mathcal{A}^{(n)} \subseteq \mathcal{X}^n$ ناحیه‌ی پذیرش فرضیه‌ی $H_۱$ باشد. احتمالات خطا را نیز به صورت زیر تعریف می‌کنیم:

$$\alpha_n = P_۱^{(n)}(\mathcal{X}^n \backslash \mathcal{A}^{(n)}), \qquad \beta_n = P_۲^{(n)}(\mathcal{A}^{(n)}).$$

در این صورت، احتمال خطای کل برابر خواهد شد با:

$$\mathbb{P}_{\mathsf{E}}^{(n)} = \pi_۱\alpha_n + \pi_۲\beta_n.$$

D^* را به صورت زیر تعریف می‌کنیم:

$$D^* = \lim_{n\to\infty} \frac{-۱}{n} \log \min_{\mathcal{A}^{(n)}\subseteq\mathcal{X}^n} \left\{\mathbb{P}_{\mathsf{E}}^{(n)}\right\}.$$

۱. نشان دهید D^* (بهترین نمای قابل دستیابی در احتمال خطای بیزی) برابر است با:

$$D^* = D_{\mathrm{KL}}(P_{\lambda^*} \| P_۱) = D_{\mathrm{KL}}(P_{\lambda^*} || P_۲),$$

که در آن داریم:

$$p_\lambda(x) = \frac{p_۱^\lambda(x) p_۲^{۱-\lambda}(x)}{\sum_{y \in \mathcal{X}} p_۱^\lambda(y) p_۲^{۱-\lambda}(y)},$$

و λ^* نیز مقداری از λ است که برای آن داریم

$$D_{\mathrm{KL}}(P_{\lambda^*} \| P_۱) = D_{\mathrm{KL}}(P_{\lambda^*} \| P_۲).$$

۲. اطّلاعات چرنف[1] بین دو توزیع $P_۱$ و $P_۲$ به صورت زیر تعریف می‌شود:

$$\mathcal{C}(P_۱, P_۲) \triangleq - \min_{۰ \le \lambda \le ۱} \log \left(\mathbb{E}_{X \sim P_۲} \left[\left(\frac{p_۱(X)}{p_۲(X)} \right)^\lambda \right] \right).$$

نشان دهید:

$$D^* = \mathcal{C}(P_۱, P_۲).$$

۵ دم‌های توزیع‌های دوجمله‌ای و پواسون

متغیّر تصادفی $X \sim \mathsf{Binomial}(n, p)$ را در نظر بگیرید.

۱. ثابت کنید:

$$\begin{aligned} \mathbb{P}[X \ge k] &\le \exp\left\{ -n \cdot D_{\mathrm{KL}}\big(\mathsf{Bernoulli}(k/n) \| \mathsf{Bernoulli}(p)\big) \right\} \quad \forall \quad k > np, \\ \mathbb{P}[X \le k] &\le \exp\left\{ -n \cdot D_{\mathrm{KL}}\big(\mathsf{Bernoulli}(k/n) \| \mathsf{Bernoulli}(p)\big) \right\}, \quad k < np \end{aligned}.$$

۲. نشان دهید که:

$$\begin{aligned} \mathbb{P}\big[X \ge u\,\mathbb{E}[X]\big] &\le \exp\{-\mathbb{E}[X] f(u)\} \forall u > ۱, \\ \mathbb{P}\big[X \le u\,\mathbb{E}[X]\big] &\le \exp\{-\mathbb{E}[X] f(u)\} \forall ۰ \le u < ۱, \end{aligned}$$

که در آن، $f(u) \triangleq u \log u - (u-۱) \ge ۰$ است.
راهنمایی: می‌توانید از نامساوی $x \log(\frac{x}{y}) \ge x - y$ که به ازای هر مقدار $x, y \in [۰, ۱]$ برقرار است، بدون اثبات استفاده کنید.

۳. نشان دهید

$$f(u) = \int_۱^u \frac{u - x}{x} dx \ge \frac{(u-۱)^۲}{۲u},$$

و نتیجه بگیرید:

$$\mathbb{P}[X > np + t] \le \exp\left(\frac{-t^۲}{۲(t + np)} \right) \quad \forall t > ۰.$$

۴. $(*)$ ثابت کنید:

$$\begin{aligned} \mathbb{P}[\sqrt{X} - \sqrt{np} \ge t] &\le e^{-t^۲}, \\ \mathbb{P}[\sqrt{X} - \sqrt{np} \le -t] &\le e^{-t^۲}, \end{aligned}$$

که در آن، $t > ۰$ فرض می‌شود.
راهنمایی: ابتدا نشان دهید $D_{\mathrm{KL}}\big(\mathsf{Bernoulli}(k/n) \| \mathsf{Bernoulli}(p)\big) \ge (\sqrt{q} - \sqrt{p})^۲$.

[1] Chernoff Information

۶ انحراف من‌درآوردی!

با توجه به فرم وردشی انحراف KL، تعمیم زیر از این انحراف ارائه شده است:

$$V_{\alpha,\beta,r,s,t}(P_X\|Q_X) = \sup_{f:\mathcal{X}\mapsto\mathbb{R}} \Big\{ \mathbb{E}_{P_X}[f(X)] - r\,\mathbb{E}_{Q_X}[f(X)] \\ - s\log\big(\mathbb{E}_{Q_X}[\exp\big(\alpha f(X)\big)]\big) - t\log\big(\mathbb{E}_{Q_X}[\exp\big(\beta f(X)\big)]\big)\Big\}$$

در اینجا (s,t) اعداد نامنفی و α,β,r اعداد حقیقی هستند.

۱. نشان دهید که چنانچه $r+\alpha s+\beta t \neq ۱$ برقرار باشد، مقدار انحراف همواره بینهایت است و در نتیجه تعریف فوق به درد نخور است!
در قسمت‌های بعدی فرض می‌کنیم که تساوی $r+\alpha s+\beta t = ۱$ برقرار است.

۲. نشان دهید که انحراف فوق همیشه نامنفی است.

۳. نشان دهید که انحراف فوق وقتی $P_X = Q_X$ باشد برابر با صفر است. آیا عکس آن درست است؟

۴. برای توزیع‌های مشترک P_{XY} و Q_{XY} نشان دهید

$$V_{\alpha,\beta,r,s,t}(P_X\|Q_X) \leq V_{\alpha,\beta,r,s,t}(P_{XY}\|Q_{XY})$$

۵. نشان دهید پردازش یکسان انحراف را افزایش نمی‌دهد، یعنی

$$V_{\alpha,\beta,r,s,t}(P_X W_{Y|X}\|Q_X W_{Y|X}) = V_{\alpha,\beta,r,s,t}(P_X\|Q_X)$$

از این‌جا نامساوی پردازش داده‌ها را برای انحراف فوق بیان کرده و ثابت نمایید. همچنین نشان دهید که انحراف فوق نسبت به زوج (P_X,Q_X) محدّب است.

۶. خاصیت بالاجمعی زیر را ثابت نمایید:

$$V_{\alpha,\beta,r,s,t}(P_{XY}\|Q_XQ_Y) \geq V_{\alpha,\beta,r,s,t}(P_X\|Q_X) + V_{\alpha,\beta,r,s,t}(P_Y\|Q_Y)$$

۷. $(*)$ قرار دهید:

$$W_\alpha(P_X\|Q_X) = V_{\alpha,۰,۱-\frac{۱}{\alpha},\frac{۱}{\alpha^۲},۰}(P_X\|Q_X).$$

حدّ زیر را بیابید:

$$\lim_{\alpha\to ۰} W_\alpha(P_X\|Q_X).$$

۸. $(*)$ مستقیماً یا با استفاده از قسمت‌های قبل نشان دهید:

$$D_{\chi^۲}(P_{XY}\|Q_XQ_Y) \geq D_{\chi^۲}(P_X\|Q_X) + D_{\chi^۲}(P_Y\|Q_Y).$$