باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

نظریه اطلاعات، آمار و یادگیری

استاد: دکتر یاسائی

نیم‌سال اول ۱۴۰۱-۱۴۰۲

تمرین سری اول

---

۱ به هم چسبیدگی!

اگر $(\mathcal{S}_n, \mathcal{F}_n)$ فضا های اندازه باشند، و $\mathbb{P} = \{\mathcal{P}_n\}_{n=1}^{\infty}, \mathbb{Q} = \{\mathcal{Q}_n\}_{n=1}^{\infty}$ دنباله ای از اندازه های احتمال روی این فضا ها باشند، می گوییم $\mathbb{P}$ به $\mathbb{Q}$ چسبیده [1] و آنرا با علامت $\mathbb{P} \lhd \mathbb{Q}$ نشان می دهیم، اگر برای هر دنباله از مجموعه ها ی $A_n \in \mathcal{S}_n$ که $\mathcal{Q}_n(A_n) \xrightarrow{n\to\infty} 0$ داشته باشیم: $\mathcal{P}_n(A_n) \xrightarrow{n\to\infty} 0$.

(۱) ثابت کنید اگر $\mathbb{P} \lhd \mathbb{Q}$ و یا $\mathbb{Q} \lhd \mathbb{P}$ در یک تست برای تشخیص بین فرضیه ی اول با توزیع $\mathcal{P}_n$ و فرضیه دوم با توزیع $\mathcal{Q}_n$ مجموع احتمال خطای نوع اول و دوم نمی تواند به صفر همگرا شود(وقتی $n \to \infty$) .

برای فضای توابع $f : \mathcal{S}_n \to \mathbb{R}$ ضرب داخلی زیر را تعریف می کنیم:

$$\langle f, g\rangle = \mathbb{E}_{X\sim\mathcal{Q}_n}[f(X)g(X)]$$

(۲) ثابت کنید اگر تحت این ضرب داخلی داشته باشیم: $\|L_n\|^2 < \infty$ (که $L_n = \frac{d\mathcal{P}_n}{d\mathcal{Q}_n}$ همان نسبت **Likelihood** است.) در این صورت داریم: $\mathbb{P} \lhd \mathbb{Q}$.

۲ انحراف بزرگ برای Log-Likelihood

فرض کنید $\mathcal{P}$ و $\mathcal{Q}$ دو توزیع احتمال باشند که $\mathcal{P} \ll \mathcal{Q}$، همینطور X_i ها متغیر های تصادفی $i.i.d.$ از $log\frac{\mathcal{P}}{\mathcal{Q}}$ تحت توزیع $\mathcal{P}$ باشند و Y_i ها متغیر های تصادفی $i.i.d.$ از $log\frac{\mathcal{P}}{\mathcal{Q}}$ تحت توزیع $\mathcal{Q}$ باشند. در این سوال می خواهیم رابطه ی زیر را ثابت کنیم:

$$\forall x \geq 0, n \in \mathbb{N} \quad \mathbb{P}[\sum_{i=1}^{n}(Y_i - X_i) \geq nx] \leq exp(-n(\alpha + \frac{x}{2}))$$

که در آن $\alpha = -2log\mathcal{B}(\mathcal{P}, \mathcal{Q})$ و $\mathcal{B}(\mathcal{P}, \mathcal{Q}) = \mathbb{E}_{\mathcal{Q}}[\sqrt{\frac{d\mathcal{P}}{d\mathcal{Q}}}]$

(۱) ابتدا ثابت کنید:

$$\mathbb{P}[\sum_{i=1}^{n}(Y_i - X_i) \geq nx] \leq exp(-nF(x))$$

که در آن $F(x) = \sup_{\theta\geq 0}\{\theta x - \psi_{\mathcal{P}}(-\theta) - \psi_{\mathcal{Q}}(\theta)\}$ و $\psi_{\mathcal{P}}(\theta) = log\mathbb{E}[e^{\theta X_1}], \psi_{\mathcal{Q}}(\theta) = log\mathbb{E}[e^{\theta Y_1}]$.

[1] Contiguous

(۲) ثابت کنید: $F(0) = -\psi_{\mathcal{P}}(-\frac{1}{2}) - \psi_{\mathcal{Q}}(\frac{1}{2}) = \alpha$

(۳) ثابت کنید: $F(x) \geq F(0) + \frac{x}{2}$. سپس حکم را نتیجه بگیرید.

۳ انحراف بزرگ برای توزیع ارلانگ

متغیر تصادفی $X \sim Eralng(n, \lambda)$ درواقع مجموع n متغیر تصادفی $i.i.d.$ با توزیع $exp(\lambda)$ است.

(۱) ثابت کنید تابع چگالی احتمال متغیر تصادفی $X \sim Eralng(n, \lambda)$ برابر است با:

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}$$

(۲) اگر $X_i \overset{i.i.d.}{\sim} exp(1)$ ثابت کنید:

$$\forall \xi > 1 \quad \mathbb{P}[\sum_{i=1}^{n} X_i \geq n\xi] \leq exp(-n(\xi - \log \xi - 1))$$

(۳) با فرض قسمت قبل ثابت کنید:

$$\forall \xi < 1 \quad \mathbb{P}[\sum_{i=1}^{n} X_i \leq n\xi] \leq exp(-n(\xi - \log \xi - 1))$$

۴ انحراف چرنف

آزمون فرض زیر را درنظر بگیرید:

$$H_0 : X_1, ..., X_n \overset{i.i.d.}{\sim} \mathcal{P}$$
$$H_1 : X_1, ..., X_n \overset{i.i.d.}{\sim} \mathcal{Q}$$

حالت بیزی با تویزیع یکنواخت روی فرض ها درنظر بگیرید. فرض کنید I اندیس فرض انتخاب شده و $\hat{I}$ تخمین ما از I با توجه به نمونه هاست. نشان دهید درحالت کلی نرخ بهینه کاهش خطا از رابطه زیر بدست می آید:

$$-\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}[I \neq \hat{I}] = \max_{0 \leq s \leq 1} \{-\log \mathbb{E}_{\mathcal{Q}}[(\frac{\mathcal{P}(X)}{\mathcal{Q}(X)})^s]\}$$

عبارت سمت راست تساوی به انحراف چرنف معروف است.

۵ برخی از خواص Total Variation

(۱) ثابت کنید

$$d_{TV}(\prod_{i=1}^{N} P_i, \prod_{i=1}^{N} Q_i) \leq \sum_{i=1}^{N} d_{TV}(P_i, Q_i)$$

(۲) ثابت کنید

$$d_{TV}(P_X, Q_X) = d_{TV}(P_{g(X)}, Q_{g(X)})$$

که در آن $g(x)$ تابعی یک به یک است.

(۳) ثابت کنید

$$d_{TV}(P_0, P_1) = d_{TV}(P_0 \otimes Q, P_1 \otimes Q)$$

(۴) ثابت کنید

$$d_{TV}(\mathcal{N}(0, \Sigma), \mathcal{N}(\theta, \Sigma)) = 1 - 2\Phi(||\theta \Sigma^{-1/2}||_2/2), \Phi(a) = \int_a^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

(راهنمایی: ابتدا مسئله را برای یک بعد حل کنید سپس با استفاده از عملیات سفید کردن و روابطی که در قسمت (۲) و (۳) ثابت کردید، جواب یک بعد را به تعداد بعد دلخواه تعمیم دهید. دقت کنید ماتریس Σ ماتریس مثبت معین هست.)

۶ نگاه بیزی و خواص $\mathcal{R}(\mathcal{P}, \mathcal{Q})$

در درس، روش کاهش خطای β به ازای حداقل احتمال موفقیت α به عنوان یک معیار برای به دست آوردن یک روش تصمیم گیری در مسئله آزمون فرض مشاهده کردیم. یک روش دیگری برای به دست آوردن یک روش تصمیم گیری در مسئله آزمون فرض به صورت زیر است

$$\min_{P(Z|X)} \pi_0\pi_{1|0} + \pi_1\pi_{0|1} \tag{۱}$$

که در آن π_0، π_1 به ترتیب احتمال اولیه فرض $H_0, X \sim \mathcal{P}$، $H_1, X \sim \mathcal{Q}$ است. به این معیار، معیار بیزی می گویند. حال فرض کنید که برای مسئله آزمون فرض منحنی مرزی پایینی ناحیه $\mathcal{R}(\mathcal{P}, \mathcal{Q})$ برابر α^2 باشد. اگر به ازای یک مقدار مشخص از احتمال های اولیه π_0، π_1 جواب مسئله بهینه سازی خطای کل در رابطه ۱، یک روش تصمیم گیری به صورت $Log - Likeloihood - Ratio(LLR)$ که در قضیه $Neyman - Pearson$ مطرح شد، بشود، که این روش به صورت زیر تعریف می شود

$$Z(x) = \begin{cases} 1 & log(\frac{P(X)}{Q(x)} \leq log(\tau) \\ 0 & log(\frac{P(X)}{Q(x)} > log(\tau)) \end{cases}$$

برای راحتی، فرض کنید $P(log(\frac{P(X)}{Q(x)} = 0) = 0$.

(۱) مقدار τ را بر حسب مقدار π_0، π_1 به دست آورید.

(۲) برای روش تصمیم گیری LLR که با معیار بیزی و توزیع اولیه فوق بهینه است، مقدار α،β را برحسب مقدار احتمال های اولیه π_0، π_1 به دست آورید.

(۳) برای اینکه جواب بهینه معیاری بیزی، بتواند به فرمت LLR، که در قضیه $Neyman - Pearson$ مطرح شد، بیان شود. چه شرایط روی احتمال های اولیه π_0، π_1 باید برقرار باشد؟ (دقت کنید که خطای اول و دوم تصمیم گیری LLR مطرح شده در قضیه $Neyman - Pearson$ باید روی منحنی مرزی پایینی ناحیه $\mathcal{R}(\mathcal{P}, \mathcal{Q})$ باشد.)

۷ زوج نرخ های قابل دسترس

در درس، دیدیم یک روش بررسی رفتار حدی خطا های مسئله آزمون فرض آن است که که خطای $\pi_{1|0}$ را کوچک نگه داریم و نرخ های همگرایی قابل دسترس برای خطای $\pi_{0|1}$ را به دست آوریم. حال در این مسئله می خواهیم برای هر دو عبارت خطا نرخ همگرایی به دست آوریم. منحنی مرزی ناحیه زوج نرخ های همگرایی قابل دسترس یعنی زوج نرخ هایی E_0 و E_1 که برای آن ها روش تصمیم گیری وجود دارد که در آن داریم

$$\pi_{1|0} \leq 2^{-nE_0}, \pi_{0|1} \leq 2^{-nE_1}$$

(۱) ابتدا استدلال کنید که چرا ناحیه $(2^{-nE1}, 2^{-nE_2})$ ، باید یک ناحیه محدب باشد؟

(۲) با استفاده از قضیه $Neyman - Pearson$ و قرار دادن پارامتر $\tau = n\theta$ ، که τ پارامتر روش تصمیم گیری LLR هست، نشان دهید

$$\pi^n_{1|0} \leq e^{-n\phi^*_P(\theta)}, \pi^n_{0|1} \leq e^{-n(\phi^*_Q(\theta))}, -D(P||Q) \leq \theta \leq D(Q||P)$$

که در رابطه فوق داریم $\phi^*_P(\theta) = \sup_{\lambda\in R} \lambda\theta - \phi_P(\lambda)$ ،$\phi_P(\lambda) = logE_P(e^{(\lambda log\frac{dP(X)}{dQ(X)})})$

(۳) با استفاده از نامساوی های فوق نشان دهید که زوج نرخ زیر قابل دسترس هستند.

$$E_0(\theta) = \phi^*_P(\theta), E_1(\theta) = \phi^*_P(\theta) + \theta$$

(۴) حال نشان دهید که منحنی پارامتری $E_0(\theta) = \phi^*_P(\theta), E_1(\theta) = \phi^*_P(\theta) + \theta$ ، همان ناحیه ی مرزی زوج های قابل دسترس است.

(۵) حال با استفاده از خط مرزی که برای ناحیه ی زوج نرخ های همگرایی قابل دسترس به دست آوردیم، مسئله به دست آوردن نرخ بهینه همگرایی عبارت

$$\min_{P(Z|X^n)} \pi_0\pi_{1|0} + \pi_1\pi_{0|1}$$

، به ازای مقادیر ثابت احتمال های اولیه، به صورت یک مسئله $maxmin$ به دست آورده و نشان دهید نرخ بهینه برابر است با $\phi^*_P(0)$