سؤال دوم پایان ترم ITSL محمد محمدیان بیشه 40110342

Subject :   Date :

۲. (آ) دقت کنید که به وضوح $\nu \ll \mu$ چون در غیر این صورت $D(\nu\|\mu) = \infty$.

ابتدا اثبات می‌کنیم که همیشه مقدار عبارت سمت راست کم‌تر مساوی است، سپس حالت تساوی ارائه می‌دهیم:

⊛ اثبات کران:

$$\mathbb{E}_\nu[Z] - D(\nu\|\mu) = \int Z\,d\nu - \int \log\left(\frac{d\nu}{d\mu}\right) d\nu =$$

$$= \int \left(Z - \log\left(\frac{d\nu}{d\mu}\right)\right) d\nu$$

$f = \frac{d\nu}{d\mu}$ را تعریف می‌کنیم

$$= \int \log\left(e^Z \times \frac{1}{f}\right) d\nu$$

طبق jensen $\longrightarrow$

$$\leq \log\left(\int \frac{e^Z}{f}\, d\nu\right) = \log\left(\int e^Z\, d\mu\right)$$

$$= \log \mathbb{E}_\mu[e^Z]$$

پس داریم:

$$\mathbb{E}_\nu[Z] - D(\nu\|\mu) \leq \log \mathbb{E}_\mu[e^Z]$$

با گرفتن supremum داریم:

$$\sup_{\nu:\, D(\nu\|\mu)<\infty} \left\{\mathbb{E}_\nu[Z] - D(\nu\|\mu)\right\} \leq \log \mathbb{E}_\mu[e^Z]$$

⊛ ایجاد حالت تساوی: تعریف می‌کنیم:

$$\frac{d\nu^*}{d\mu} := \frac{e^Z}{\mathbb{E}_\mu[e^Z]}$$

$$\longrightarrow D(\nu^*\|\mu) = \int \log\left(\frac{e^Z}{\mathbb{E}_\mu[e^Z]}\right) d\nu^* = \mathbb{E}_{\nu^*}[Z] - \log \mathbb{E}_\mu[e^Z]$$

$$\Longrightarrow \log \mathbb{E}_\mu[e^Z] = \mathbb{E}_{\nu^*}[Z] - D(\nu^*\|\mu)$$

MICRO

پس حالت تساوی هم رخ می دهد. پس داریم:

$$\log \mathbb{E}_{\mu}[e^{z}] = \sup_{v:\, D(v\|\mu)<\infty} \left\{ \mathbb{E}_{v}[z] - D(v\|\mu) \right\} \qquad \blacksquare$$

ب) برای راحتی کار در این بخش، تعریف می کنیم:

$$M(\theta) := \mathbb{E}_{x}\left[e^{f(x,\theta)}\right]$$

تعریف می کنیم:

$$g_x(\theta) := f(x,\theta) - \log M(\theta)$$

دقت کنید که:

$$\mathbb{E}_{x}\left[e^{g_x(\theta)}\right] = \frac{\mathbb{E}_{x}\left[e^{f(x,\theta)}\right]}{M(\theta)} = 1$$

دقت کنید که عبارت معادل را در جلو با g می توان بهتر تعریف کرد.

$$\mathbb{E}_{v}[f(x,\theta)] - \mathbb{E}_{v}[\log M(\theta)] = \mathbb{E}_{v}(g_x(\theta))$$

پس واقعه موجود در جلو معادل است با:

$$\left\{ \exists_{v} :\ \mathbb{E}_{v}(g_x(\theta)) - D(v\|\mu) \geq t \right\}$$

این واقعه را به صورت معادل تعریف می کنیم:

$$\left\{ \sup_{v} \left(\mathbb{E}_{v}(g_x(\theta)) - D(v\|\mu) \right) \geq t \right\}$$

حال تعریف می کنیم:

$$Y := \exp\left(\sup_{v} \left\{ \mathbb{E}_{v}[g_x(\theta)] - D(v\|\mu) \right\} \right)$$

طبق نامساوی مارکف داریم:

$$(*) \quad \boxed{P_x(Y \geq e^{t}) \leq e^{-t}\,\mathbb{E}_{x}[Y] = e^{-t}\,\mathbb{E}_{x}\left[\exp\left(\sup_{v} \left\{ \mathbb{E}_{v}[g_x(\theta)] - D(v\|\mu) \right\} \right) \right]}$$

پس برای اثبات حکم کافی است اثبات کنیم:

$$\mathbb{E}_X\left[\exp\left(\sup_{\nu}\{\mathbb{E}_\nu[g_X(\theta)]-D(\nu\|\mu)\}\right)\right] \le 1$$

از بخش (آ) همین سوال می‌دانیم: $\mathbb{E}_\mu[e^{g_X(\theta)}] = \exp\left(\sup_{\nu}\{\mathbb{E}_\nu[g_X(\theta)]-D(\nu\|\mu)\}\right)$

پس داریم

$$\mathbb{E}_X\left[\exp\left(\sup_{\nu}\{\mathbb{E}_\nu[g_X(\theta)]-D(\nu\|\mu)\}\right)\right] = \mathbb{E}_X\mathbb{E}_\mu[e^{g_X(\theta)}]$$

$$= \mathbb{E}_\mu\mathbb{E}_X[e^{g_X(\theta)}]$$

$$= \mathbb{E}_\mu[1]$$

$$= 1 \quad .$$

پس طبق $(*)$ و تساوی بالا:

$$P_X\left(\sup_{\nu}\{\mathbb{E}_\nu[g_X(\theta)]-D(\nu\|\mu)\}\ge t\right) \le e^{-t}\cdot 1 = e^{-t}$$

که همان حکم است. ■

ج) طبق راهنمایی: $V_{uv} := \mathrm{Unif}(\{a : \|a-u\| \le \varepsilon\}) \times \mathrm{Unif}(\{b : \|b-v\| \le \varepsilon\})$

که روی Θ است.

برای محاسبه $D(V_{uv}\|\mu)$، با توجه به اینکه داریم: $D(\mathrm{unif}(A)\|\mathrm{unif}(B)) = \log \frac{Vol(A)}{Vol(B)}$

از طرفی $Vol(rB^d) = r^d Vol(B^d)$ پس داریم:

$$D(V_{uv}\|\mu) = \log \frac{Vol((1+\varepsilon)B^m)}{Vol(\varepsilon B^m)} + \log \frac{Vol((1+\varepsilon)B^n)}{Vol(\varepsilon B^n)}$$

$$= m \log \frac{1+\varepsilon}{\varepsilon} + n \log \frac{1+\varepsilon}{\varepsilon} = (m+n) \log \frac{1+\varepsilon}{\varepsilon}$$

حال که مقدار $D(V_{uv}\|\mu)$ را بدست آوردیم، f را این‌طور تعریف می‌کنیم: $(a,b)\in\Theta$ $f(X;(a,b)) := \lambda a^T X b$

دقت کنید که $a^T X b = \sum_{i,j} a_i b_j X_{ij}$.

از طرفی میانگین این متغیر (با توجه به گاوسی نرمال بودن X_{ij} ها) صفر است، و واریانس آن برابر است با:

$$Var(a^T X b) = \sum_{i,j} a_i^2 b_j^2 = \left(\sum a_i^2\right)\left(\sum b_j^2\right) = \|a\|^2 \|b\|^2$$

پس داریم: $$\log E_X[e^{f(X;(a,b))}] = \log E_X[e^{\lambda a^T X b}] = \frac{\lambda^2}{2} \|a\|^2 \|b\|^2$$

حال چون $(a,b) \in \Theta$ پس $\|a\| \le 1+\varepsilon$ و $\|b\| \le 1+\varepsilon$ پس داریم:

$$E_{V_{uv}}[\log E_X[e^{f(X;(a,b))}]] \le \frac{\lambda^2}{2}(1+\varepsilon)^4$$

از طرفی برای مقدار $\mathbb{E}_{\nu_{u,v}}[f(X,\theta)]$ داریم:

چون $\mathbb{E}_{\nu_{u,v}}[a] = u$, $\mathbb{E}_{\nu_{u,v}}[b] = v$

پس برای X ثابت: $\mathbb{E}_{\nu_{u,v}}[f(X,\theta)] = \lambda\, \mathbb{E}[a]^T X\, \mathbb{E}[b] = \lambda u^T X v$

حال دقت کنید که طبق بخش (ب) داریم:

بخش ب $\rightarrow$ $P_X[\exists v: \mathbb{E}_v[f(X,\theta)] \geq \mathbb{E}_v[\log \mathbb{E}_X[\exp(f(X,\theta))]] + D(v\|\mu) + t] \leq e^{-t}$

پس $\Longrightarrow$ $\boxed{P_X(\forall v: \mathbb{E}_v[f(X,\theta)] \leq \mathbb{E}_v[\log \mathbb{E}_X[\exp(f(X,\theta))]] + D(v\|\mu) + t) \geq 1 - e^{-t}}$ (*)

حال با تعریف: $f(X;(a,b)) = \lambda a^T X b$ و $\mu := Unif((1+\epsilon)B^m) \times Unif((1+\epsilon)B^n)$

و با توجه به تعریف $\nu_{u,v}$، این مقادیر را در معادله بدست آمده در بخش (ب) قرار دهیم، داریم:

قرار دادن مقادیر در (*) $\rightarrow$ عبارت درون P_X $\rightarrow$ $\lambda u^T X v \leq \frac{\lambda^2}{2}(1+\epsilon)^4 + (m+n)\log\frac{1+\epsilon}{\epsilon} + t$

با توجه به استفاده از نامساوی $\frac{\lambda^2}{2}(1+\epsilon)^4 \geq \mathbb{E}_{\nu_{u,v}}[\log \mathbb{E}_X[e^{f(X;(a,b))}]]$ در خط بالا می‌توانیم نتیجه بگیریم که پیشامد (*) را می‌توان

به صورت مقابل نشان داد: $\|X\| = \sup_{u,v} u^T X v \leq \frac{\lambda}{2}(1+\epsilon)^4 + \frac{(m+n)\log\frac{1+\epsilon}{\epsilon} + t}{\lambda}$

حال کافی است این پیشامد را به فرم حکم سوال بدست آورد:

فرم سمت راست نامساوی به فرم $A\lambda + \frac{B}{\lambda}$ است. مینیمم این تابع نسبت به λ، چون λ مثبت است، برابر است با:

$$f(A,B) = A\lambda + \frac{B}{\lambda} \longrightarrow \frac{d}{d\lambda} f(A,B) = A - \frac{B}{\lambda^2} = 0 \rightarrow \lambda^* = \sqrt{\frac{B}{A}}$$

$$\longrightarrow f^*(A,B) = 2\sqrt{AB}$$

پس می‌توان سمت راست پیشامد احتمالی را به صورت زیر نوشت:

$$\|X\| \leq \sqrt{2}\,(1+\varepsilon)^2 \sqrt{(m+n)\log\frac{1+\varepsilon}{\varepsilon} + t}$$

حال با قرار دادن $\varepsilon = 1$ داریم، $\log\frac{1+\varepsilon}{\varepsilon} = \log 2 \leq 1$ پس

$$\|X\| \leq \sqrt{2}\,(1+1)^2 \sqrt{m+n+t}$$

پس:

$$\|X\| \leq 4\sqrt{2}\sqrt{m+n+t}$$

پس $C = 4\sqrt{2}$.

پس ما ثابت کردیم که:

$$P\left(\|X\| \leq C\sqrt{m+n+t}\right) \geq 1 - e^{-t}$$

که در آن $t > 0$, $C = 4\sqrt{2}$. ∎