Subject : Date :

۱. خواص Total Variation

۱. برای اثبات حکم این بخش، روی n استقرا می‌زنیم. ابتدا برای حالت $n=2$ حکم را اثبات می‌کنیم:

برای حالت $n=2$، فرض کنید دو متغیر تصادفی $X_1 \in \{x_1, \dots, x_n\}$ و $X_2 \in \{y_1, \dots, y_m\}$ داریم که X_1 با توزیع‌های p_1 و q_1 و X_2 با توزیع‌های p_2 و q_2 تولید شونده داریم:

$$d_{TV}(p_1 p_2, q_1 q_2) \leq d_{TV}(p_1, q_1) + d_{TV}(p_2, q_2)$$

$$\iff \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i}\, p_{2,j} - q_{1,i}\, q_{2,j}| \leq \frac{1}{2}\sum_{i=1}^{n} |p_{1,i} - q_{1,i}| + \frac{1}{2}\sum_{j=1}^{m} |p_{2,j} - q_{2,j}|$$

که به عنوان مثال $p_{1,i}$ احتمال $P_1(X_1 = x_i)$ است یا $q_{2,j}$ همان $Q_2(X_2 = x_j)$ است.

برای اثبات حکم، از محدب بودن قدرمطلق استفاده می‌کنیم؛ با توجه به نمسن دون دهی شده داریم: $|ab - cd| \leq |a-c|b + |b-d|c$

با توجه به این نامساوی داریم:

$$|p_{1,i}\, p_{2,j} - q_{1,i}\, q_{2,j}| \leq |p_{1,i} - q_{1,i}|\, p_{2,j} + q_{1,i}\, |p_{2,j} - q_{2,j}|$$

جمع می‌کنیم

$$\sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i}\, p_{2,j} - q_{1,i}\, q_{2,j}| \leq \sum_{i=1}^{n}\sum_{j=1}^{m} |p_{1,i} - q_{1,i}|\, p_{2,j} + q_{1,i}\, |p_{2,j} - q_{2,j}| \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} |p_{1,i} - q_{1,i}| \underbrace{\sum_{j=1}^{m} p_{2,j}}_{1} + \underbrace{\sum_{i=1}^{n} q_{1,i}}_{1} \sum |p_{2,j} - q_{2,j}|$$

$$= \sum_{i=1}^{n} |p_{1,i} - q_{1,i}| + \sum_{j=1}^{m} |p_{2,j} - q_{2,j}|$$

پس حکم در حالت $n=2$ اثبات می‌شود.

Subject : Date :

فرض و گام استقرا: فرض کنید حکم مسئله برای $n=k$ درست باشد، حال برای $n=k+1$ داریم:

$$\sum_{i=1}^{k+1} d_{TV}(p_i,q_i) = \sum_{i=1}^{k} d_{TV}(p_i,q_i) + d_{TV}(p_{k+1},q_{k+1}) \geq d_{TV}\left(\prod_{i=1}^{k} p_i, \prod_{i=1}^{k} q_i\right) + d_{TV}(p_{k+1},q_{k+1})$$

فرض استقرا

حال برای نامساوی اخیر، با توجه به $n=2$ به سادگی حکم زیر برقرار است:

$$d_{TV}\left(\prod_{i=1}^{k} p_i, \prod_{i=1}^{k} q_i\right) + d_{TV}(p_{k+1},q_{k+1}) \geq d_{TV}\left(\prod_{i=1}^{k+1} p_i, \prod_{i=1}^{k+1} q_i\right)$$

پس حکم مسئله برای $n=k+1$ نیز اثبات می‌شود و حکم مسئله اثبات شد.

Subject :   Date :

(الف) دقت کنید که داریم:

$$L := \frac{dP}{dQ} \; (\geq 0) \quad , \quad \mathbb{E}_Q[L] = 1$$

$$\to \; TV(P,Q) = \frac{1}{2}\mathbb{E}_Q[|L-1|]$$

$$\to \; \chi^2(P \| Q) = \mathbb{E}_Q[(L-1)^2]$$

حال که ما TV و χ^2 را به صورت تابعی از L تعریف کردیم، قصد تحلیل و به گونه‌ای "فشرده" کردن L را داریم، هدف از اینجا، اثبات Proposition 7.15 در کتاب Polyanskiy به گونه ذیل است:

تعریف می‌کنیم:

$$A := \{L \geq 1\} \quad , \quad B := \{L < 1\} \quad , \quad \alpha := Q(A)$$

همچنین:

$$a := \mathbb{E}[L|A] \to (\geq 1) \qquad b := \mathbb{E}[L|B] \; (< 1)$$

حال L را به این صورت "فشرده" تعریف می‌کنیم:

$$\tilde{L} := a\mathbb{1}_A + b\mathbb{1}_B$$

دقت کنید با توجه به تعریف a و b، A، B، $\tilde{L}$ به گونه‌ای همان میانگین L است، و فشرده شده likelihood آن است؛ داریم:

$$\mathbb{E}_Q[|\tilde{L}-1|] = \alpha(a-1) + (1-\alpha)(1-b) = \mathbb{E}_Q[|L-1|]$$

پس با فشرده سازی، TV تغییری نمی‌کند اما برای χ^2 داریم:

$$E_Q[(\bar{L}-1)^2] \leqslant E_Q[(L-1)^2] = \chi^2$$

دقت کنید که نامساوی اخیر، با توجه به نامساوی Jensen و محدب بودن تابع $f(x)=(x-1)^2$ بدیهی است.

پس برای TV ثابت، بالاترین مقداری که می‌توان آن را به χ^2 ربط داد، زمانی که L دو مقداری باشد رخ می‌دهد. حال L را به صورت دو مقداری تعریف می‌کنیم:

$$L = \begin{cases} 1+v & \xrightarrow{\in A} \quad Q(A)=\alpha \\ 1-u & \xrightarrow{\in B} \quad Q(B)=1-\alpha \end{cases} \quad \rightarrow u\in[0,1], \quad v\geqslant 0$$

دقت کنید که مقادیر $1+v$ و $1-u$ به دلیل $\mathbb{E}[L|A]\geqslant 1$ و $\mathbb{E}[L|B]<1$ تعریف کردیم.

حال سعی در محاسبه u و v برحسب TV داریم:

$$\left.\begin{aligned} 2TV &= \mathbb{E}[|L-1|] = \alpha v + (1-\alpha)u \\ 1 &= \mathbb{E}[L] = \alpha(1+v) + (1-\alpha)(1-u) \Rightarrow \alpha v = (1-\alpha)u \end{aligned}\right\} \quad v = \frac{TV}{\alpha}, \quad u = \frac{TV}{1-\alpha}$$

حال برای χ^2 داریم:

$$\chi^2 = \mathbb{E}[(L-1)^2] = \alpha v^2 + (1-\alpha)u^2 = (TV)^2\left(\frac{1}{\alpha}+\frac{1}{1-\alpha}\right) = \frac{TV^2}{\alpha(1-\alpha)}$$

Subject : Date :

$$u = \frac{S}{1-\alpha}$$

حال برای راحتی کار تعریف می کنیم $TV = S$ داریم : چون $L \geq 0 \Rightarrow u \leq 1 \Rightarrow \alpha \leq 1-S$

دقت کنید داریم :
$$\chi^2 = \frac{S^2}{\alpha(1-\alpha)}$$

حال برای $TV = S$ ثابت، برای کمینه کردن χ^2 باید مقدار $\alpha(1-\alpha)$ را بیشینه کنیم، دو حالت داریم :

① اگر این TV ثابت در بازه $0 \leq S \leq \frac{1}{2}$ باشد :

در این حالت α می تواند $\frac{1}{2}$ باشد و به راحتی داریم :
$$\chi^2_{min}(S) = \frac{S^2}{\frac{1}{2}(\frac{1}{2})} = 4S^2$$

② اگر $1 > S \geq \frac{1}{2}$ باشد، با توجه به $\alpha \leq 1-S$، α می تواند ماکسیمم کل خود را اتخاذ کند و $\frac{1}{2}$ باشد و ماکسیمم در این حالت برای χ^2 زمانی رخ می دهد که $\alpha = 1-S$ باشد :
$$\chi^2_{min}(S) = \frac{S^2}{(1-S)S} = \frac{S}{1-S}$$

پس داریم :
$$\chi_{min} = \begin{cases} 4S^2 & 0 \leq S \leq \frac{1}{2} \\ \frac{S}{1-S} & \frac{1}{2} \leq S < 1 \end{cases}$$

پس داریم :
$$\chi^2(P\|Q) \geq \chi^2_{min}(TV(P,Q))$$

پس Proposition 7.15 اثبات شد.

حال دقت کنید تابع $g(x) = \sqrt{\frac{x}{1+x}}$ برای $x \geq 0$ تابعی صعودی است. برای راحتی قرار می دهیم $t = \chi^2(P\|Q)$

داریم:

$$g(t) \geq g(\chi^2_{\min}(TV(P,Q)))$$

① برای $0 \leq S \leq \frac{1}{2}$ داریم:

$$g(4S^2) = \frac{2S}{\sqrt{1+4S^2}} \geq S \longrightarrow \left(\sqrt{1+4S^2} \leq 2 \text{ for } [0, \tfrac{1}{2}]\right)$$

② برای $\frac{1}{2} \leq S \leq 1$:

$$g\left(\frac{S}{1-S}\right) = \sqrt{S} \geq S$$

پس به طور کلی: $g(t) \geq S$ پس:

$$\sqrt{\frac{\chi^2(P\|Q)}{1+\chi^2(P\|Q)}} \geq TV(P,Q)$$

پس حکم برای $c=1$ برقرار است.

حال کافی است ثابت کنیم $c=1$ بهینه‌ترین حالت است. برای اینکار مثالی Bernolli ارائه می‌دهیم که به $c=1$ برسد:

دو توزیع P, Q را Binary درنظر بگیرید با احتمال‌های: $P \Rightarrow \{1, 0\}$, $Q \Rightarrow (\varepsilon, 1-\varepsilon)$

داریم:

$$TV(P,Q) = 1-\varepsilon \quad , \quad \chi^2(P\|Q) = \frac{(1-\varepsilon)^2}{\varepsilon} + 1-\varepsilon = \frac{1-\varepsilon}{\varepsilon}$$

حال دقت کنید:

$$\frac{TV}{\sqrt{\frac{\chi^2}{1+\chi^2}}} = \frac{1-\varepsilon}{\sqrt{\frac{\frac{1-\varepsilon}{\varepsilon}}{1+\frac{1-\varepsilon}{\varepsilon}}}} = \frac{1}{\sqrt{1-\varepsilon}}$$

Subject : Date :

حال دقت کنید که برای $\lim_{\varepsilon \to 0}$ داریم:

$$\lim_{\varepsilon \to 0} \frac{TV}{\sqrt{\frac{x^2}{1+x^2}}} = \lim_{\varepsilon \to 0} \frac{1}{\sqrt{1-\varepsilon}} = 1$$

پس برای هر c ای با شرط $c<1$ ، مثال نقضی با توجه به معادله بالا ارائه دادیم. پس c بهینه همان 1 است.

$$\boxed{c^* = 1}$$

و حکم به طور کامل اثبات می‌شود. ■

Subject : Date :

سوال ۱ (ج) TV بین دو توزیع نرمال گوسی

ابتدا راهنمایی داده شده را اثبات می کنیم:

دقت کنید با تعریف C^t، می دانیم که $C^0 = I$ ، $C^1 = C$

از طرفی داریم: $\log(C) = \sum_{i=1}^{d} \ln c_i u_i u_i^T$. با توجه به این تعریف، برای مشتق های C^t ، C^{-t} نسبت به t خواهیم داشت:

$$\frac{d}{dt}C^t = C^t \log C \quad \leftarrow \quad \left(\text{چون } \frac{d}{dt}c_i^t = c_i^t \ln c_i\right)$$

$$\frac{d}{dt}(C^t)^{-1} = -(C^t)^{-1}\frac{d}{dt}C^t(C^t)^{-1} = -C^{-t}\log C$$

حال با توجه به تعریف f_B در راهنمایی، داریم: $f_B(x) = (2\pi)^{-\frac{d}{2}}(\det B)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}x^T B x\right)$

حال دقت کنید که خروجی $f_B(x)$ یک عدد است (احتمال). حال قرار می دهیم $B = C^t$ داریم:

$$\log f_{C^t}(x) = const - \frac{1}{2}\log\det(C^t) - \frac{1}{2}x^T C^t x \qquad (\text{Equation 1})$$

برای مشتق گیری از عبارت بالا، طبق خواص ماتریس ها داریم که: $\frac{d}{dt}\log\det\Sigma(t) = tr\left(\Sigma'(t)\Sigma(t)^{-1}\right)$

و همچنین، $\frac{d}{dt}\Sigma(t)^{-1} = -\Sigma(t)^{-1}\Sigma'(t)\Sigma(t)^{-1}$

حال با قرار دادن $\Sigma(t) = C^t$ داریم:

Subject : Date :

$$\frac{d}{dt}\log\det C^t = tr(C^{t}\log C\, C^{-t}) = tr(\log C)$$

$$\frac{d}{dt}(x^T C^{-t} x) = x^T\left(\frac{d}{dt}C^{-t}\right)x = -x^T C^{-t}(C^t \log C)C^{-t}x = -x^T \log(C)\, C^{-t} x$$

حال با مشتق گیری از Equation 1 داریم:

$$\frac{d}{dt}\log f_{C^t}(x) = -\frac{1}{2}tr(\log C) + \frac{1}{2}x^T(\log C)C^{-t}x$$

از طرفی با توجه به: $\frac{d}{dt}f_{C^t}(x) = f_{C^t}(x)\frac{d}{dt}\log f_{C^t}(x)$ داریم:

$$\frac{d}{dt}f_{C^t}(x) = f_{C^t}(x)\cdot\frac{1}{2}\left(x^T\log(C)C^{-t}x - tr(\log C)\right) \quad (*)$$

پس $f_{C^t}(x)$ تابعی مشتق‌پذیر از t است. و می‌توانیم از قضیه اساسی حساب استفاده کنیم:

تابع g را در نظر بگیرید که به صورت $t \rightarrow f_{C^t}(x)$ است. با توجه به مشتق‌پذیری g داریم:

$$g(b) - g(a) = \int_b^a \frac{d}{dt}(g(x))\,dx$$

با قرار دادن $a=0$ و $b=1$ داریم:

$$f_{C^1}(x) - f_{C^0}(x) = f_C(x) - f_I(x) = \int_0^1 \frac{d}{dt}f_{C^t}(x)\,dt$$

Subject : Date :

پس داریم: $f_C(x) - f_I(x) = \int_0^1 \frac{d}{dt} f_{C^t}(x)\, dt$

به وضوح از نامساوی مثلث خواهیم داشت:

$$|f_C(x) - f_I(x)| \le \int_0^1 \left|\frac{d}{dt} f_{C^t}(x)\right| dt$$

پس راهنمایی اثبات شد. ■

حال برای اثبات حکم مسئله داریم که: $TV(P,Q) = \frac{1}{2} \int_{\mathbb{R}^d} |f_P - f_Q|$

با توجه به راهنمایی داریم:

$$(**)\ \boxed{TV(N(0,C), N(0,I)) \le \frac{1}{2} \int_0^1 \int_{\mathbb{R}^d} \left|\frac{d}{dt} f_{C^t}(x)\right| dx\, dt}$$

با توجه به نامساوی خط بالا، حکم معادل است با اثبات: $\int \left|\frac{d}{dt} f_{C^t}(x)\right| dx \le 2\sqrt{2\sum_{i=1}^{d}(\ln c_i)^2}$

داریم:

$$\int \left|\frac{d}{dt} f_{C^t}(x)\right| dx \overset{\text{طبق } (*)}{=} E\left[\frac{1}{2}\left|\overbrace{X_t^T \log C\, C^{-t} X_t - tr(\log C)}^{Z_t}\right|\right]$$

که در آن $X_t \sim N(0, C^t)$، از طرفی با توجه به اینکه $E[X_t^T A X_t] = tr(AC^t)$ داریم:

$$E[X_t^T \log C\, C^{-t} X_t] = tr(\log C)$$

از طرفی با توجه به کوشی شوارتز داریم: $E\left[\frac{1}{2}|Z_t|\right] \le \frac{1}{2}\sqrt{E[Z_t^2]} = \frac{1}{2}\sqrt{Var(Z_t)}$

دقت کنید که :

$$Z_t = X_t^T \log C \, C^{-t} X_t - tr(\log C)$$

برای $Var(Z_t)$ با توجه به $Var(X^T A X) = 2\,tr((A\Sigma)^2)$ که برای متغیرهای گوسی درست است، داریم :

$$A \to \log C \, C^{-t}$$
$$\Sigma \to C^{t} \longrightarrow A\Sigma = \log C$$

پس :

$$Var(Z_t) = 2\,tr((\log C)^2) = 2\sum_{i=1}^{d} \ln(c_i)^2$$

پس داریم :

$$\int \left| \frac{d}{dt} f_t(x) \right| dx \leq \frac{1}{2}\sqrt{2\sum_{i=1}^{d} (\ln c_i)^2}$$

حال با قرار دادن این نامساوی در (**) داریم :

$$TV(\mathcal{N}(0,C), \mathcal{N}(0,I_d)) \leq \frac{1}{2}\int_0^1 \int_{\mathbb{R}^d} \left| \frac{d}{dt} f_{C^t}(x) \right| dx\, dt$$

$$\leq \frac{1}{2} \times 1 \times \frac{1}{2}\sqrt{2\sum_{i=1}^{d} (\ln c_i)^2} = \frac{1}{4}\sqrt{2\sum_{i=1}^{d} (\ln c_i)^2}$$