401110342 سوال سوم پایان ترم - محمد محمدیان بسید

Subject : Date :

سوال ۳. تست استقلال و آزمون فرض مرکب

Problem Overview

ابتدا قبل از پرداختن به سوالات، کمی خود مسئله را بررسی می‌کنیم:

$X \sim N(0,1)$, $Y \in \mathbb{R}^d$

$Y | X = x \sim N(\rho x, I_d - \rho\rho^T)$

چون داریم که: $Var(Y|X) = Var(Y) - Cov(Y,X)\,Var(X)^{-1}\,Cov(X,Y)$

پس می‌توان نتیجه گرفت که: $Y \sim N(0, I_d)$

پس در صورت کلی ماتریس Joint Cov این دو متغیر: $\Sigma_\rho = \begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{bmatrix}$

$\Sigma_\rho = \begin{bmatrix} 1 & \rho^T \\ \rho & I_d \end{bmatrix}$

خواسته‌ی ما این است که بررسی کنیم که بردار همبستگی ρ برابر صفر است یا خیر.

MICRO

الف) برای حل این بخش ابتدا لم را اثبات می‌کنیم:

لم ۱: اگر α خطای نوع اول و β خطای نوع دوم باشد، داریم: $\alpha+\beta \geq 1 - TV(P_0^{\otimes n}, P_1^{\otimes n})$

اثبات لم ۱: با توجه به تعریف خطاها داریم: $\alpha = \mathbb{E}_{P_0^{\otimes n}}[\varphi] \quad , \quad \beta = \mathbb{E}_{P_1^{\otimes n}}[1-\varphi]$

$$\rightarrow \alpha+\beta = 1 - \left(\mathbb{E}_{P_1^{\otimes n}}[\varphi] - \mathbb{E}_{P_0^{\otimes n}}[\varphi]\right)$$

داریم:

$$\mathbb{E}_{P_1^{\otimes n}}[\varphi] - \mathbb{E}_{P_0^{\otimes n}}[\varphi] = \int \varphi(x)\left(dP_1^{\otimes n}(x) - dP_0^{\otimes n}(x)\right)$$

دقت کنید که $0 \leq \varphi(x) \leq 1$ پس داریم:

$$\int \varphi(x)\left(dP_1^{\otimes n}(x) - dP_0^{\otimes n}(x)\right) \leq \sup_{0\leq\varphi\leq 1} \int \varphi(x)\left(dP_1^{\otimes n} - dP_0^{\otimes n}\right)$$

دقت کنید که بیشترین φ زمانی رخ می‌دهد که وقتی $dP_1^{\otimes n} > dP_0^{\otimes n}$ باشد، $\varphi = 1$ و در غیر این صورت صفر باشد که همان تعریف TV است:

$$\sup_{0\leq\varphi\leq 1} \int \varphi(x)\left(dP_1^{\otimes n} - dP_0^{\otimes n}\right) = TV(P_1^{\otimes n}, P_0^{\otimes n})$$

پس

$$\mathbb{E}_{P_1^{\otimes n}}[\varphi] - \mathbb{E}_{P_0^{\otimes n}}[\varphi] \leq TV(P_1^{\otimes n}, P_0^{\otimes n})$$

پس:

$$\alpha+\beta = 1 - \left(\mathbb{E}_{P_1^{\otimes n}}[\varphi] - \mathbb{E}_{P_0^{\otimes n}}[\varphi]\right) \geq 1 - TV(P_1^{\otimes n}, P_0^{\otimes n})$$

پس لم اثبات شد. ■

حال با توجه به اینکه گفته شده و خواسته مسئله: $\alpha \leq \frac{1}{10}$ ، $\beta \leq \frac{1}{10}$ داریم:

$$\frac{2}{10} \geq \alpha + \beta \geq 1 - TV(P_0^{\otimes n}, P_\rho^{\otimes n}) \Rightarrow TV(P_0^{\otimes n}, P_\rho^{\otimes n}) \geq \frac{8}{10}$$

از طرفی با توجه به نامساوی Pinsker و خواص KL داریم:

$$TV(P_0^{\otimes n}, P_\rho^{\otimes n}) \leq \sqrt{\frac{1}{2} D_{KL}(P_\rho^{\otimes n} \| P_0^{\otimes n})} = \sqrt{\frac{n}{2} D_{KL}(P_\rho \| P_0)}$$

از طرفی می‌دانیم که در P_0 داریم: $\rho = 0$ و در P_ρ داریم ~~آنگاه~~ و ماتریس joint $\Sigma_\rho = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ خواهد بود. پس P_0 و P_ρ

$P_\rho \sim N(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix})$ ، $P_0 \sim N(0, I_2)$ داریم و می‌توانیم KL آن‌ها را محاسبه کنیم:

$$D_{KL}(P_\rho \| P_0) = \frac{1}{2}\Big(\underbrace{tr(I_2^{-1}\Sigma_\rho)}_{\cancel{2}} - \log \underbrace{det(I_2^{-1}\Sigma_\rho)}_{1-\rho^2} \cancel{- 2}\Big) = -\frac{1}{2}\log(1-\rho^2)$$

پس با اعمال همه معادلات بالا داریم:

$$\frac{8}{10} \leq TV(P_0^{\otimes n}, P_\rho^{\otimes n}) \leq \sqrt{\frac{n}{2}\left(-\frac{1}{2}\log(1-\rho^2)\right)}$$

$$\Rightarrow \sqrt{\frac{n}{2}\left(-\frac{1}{2}\log(1-\rho^2)\right)} \geq \frac{8}{10} \longrightarrow n \geq \frac{2.56}{-\log(1-\rho^2)} \approx \frac{2.56}{\rho^2}$$

پس $n = \Omega(\frac{1}{\rho^2})$ و از مرتبه $\frac{1}{\rho^2}$ است.

حال برای تستی که به خواسته مسئله برسیم، تعریف می‌کنیم:
$$\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

حال دقت کنید که
$$\mathbb{E}_\rho[\hat{\rho}] = \rho \quad , \quad Var_\rho(\hat{\rho}) = \frac{1+\rho^2}{n} \leq \frac{2}{n} \quad (\rho^2 \leq 1 \text{ چون})$$

چون به طور مجزا $\mathbb{E}\{XY\}=\rho$ برای X و Y که نرمال هستند، برقرار است.

حال کافی است تست را به صورت مقابل تعریف کنیم:
$$\psi = \mathbb{I}\{\hat{\rho} \geq \frac{\tau}{2}\}$$

حال کافی است خطاهای نوع اول و دوم را محاسبه کنیم:

← خطای نوع اول:
$$P_0(\hat{\rho} \geq \frac{\tau}{2}) \underset{\text{چبیشف}}{\leq} \frac{Var_0(\hat{\rho})}{(\frac{\tau}{2})^2} = \frac{\frac{1}{n}}{\frac{\tau^2}{4}} = \frac{4}{n\tau^2}$$

که برای $n \geq \frac{40}{\tau^2}$ خواهیم داشت: $P_0\{H_1\} \leq \frac{1}{10}$

← خطای نوع دوم:
$$P_\tau(\hat{\rho} < \frac{\tau}{2}) = P_\tau(|\hat{\rho}-\tau| \geq \frac{\tau}{2}) \leq \frac{Var_\tau(\hat{\rho})}{(\frac{\tau}{2})^2} \leq \frac{\frac{2}{n}}{\frac{\tau^2}{4}} = \frac{8}{\tau^2 n}$$

که برای $n \geq \frac{80}{\tau^2}$ این نوع خطا هم کمتر از $\frac{1}{10}$ خواهد بود.

پس برای این تست، از مرتبه $\frac{1}{\tau^2}$ نمونه لازم است. ∎

Subject : Date :

ب) طبق صفحه اول حل این سوال Problem Overview، در این حالت برای Σ_ρ خواهیم داشت:

$$\Sigma_\rho = \begin{bmatrix} 1 & \rho^T \\ \rho & I_d \end{bmatrix} = \begin{bmatrix} 1 & \rho & 0 & \cdots & 0 \\ \rho & 1 & 0 & & \vdots \\ 0 & 0 & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \end{bmatrix}$$

پس $(Y_d, \dots, Y_3, Y_2) \perp\!\!\!\perp (X, Y_1)$ پس عملا $Y_2, Y_3, \dots, Y_d$ از دید H_ρ و H_0 یکی هستند.

از طرفی از دید Liklihood هم term های مربوط به این متغیرها از دو طرف ساده می‌شوند.

از دید تست هم کافی است تعریف کنیم $\hat{\rho} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_{i,1}$ و به همان نتایج قسمت قبل برسیم.

پس به طور کلی با پرسش درباره فرمول‌ها به همان حالت قبل تحلیل یافته و می‌رسیم که به تعداد نمونه در مرتبه $\frac{1}{\rho^2}$ نیاز هست. ▨

Subject :   Date :

ج) حکم این بخش، شبیه به دیگر دو ما برای بخش (آ) همین مسئله و استفاده از لم ۱ است. با این تفاوت که در بخش (آ)، p ما ثابت بود اما در اینجا p خود از توزیع π می‌آید.

فرض کنید که p را از توزیع π طبق گفته مسئله انتخاب کردیم. حال خطاهای نوع اول و دوم را به صورت α و β_p تعریف می‌کنیم.

تعریف می‌کنیم: $\bar{\beta} := E_{p\sim\pi}[\beta_p]$ ، همچنین برای راحتی کار تعریف می‌کنیم: $\bar{P} := E_{p\sim\pi}[P_p^{\otimes n}]$

دقت کنید که $\bar{\beta} = E_{p\sim\pi} P_p^{\otimes n}(\varphi=0) = \bar{P}(\varphi=0) = 1 - E_{\bar{P}}[\varphi]$

حال با توجه به فرض، تست موفق ما برای هر $p \in B_\epsilon^c$ می‌دهد: $\beta_p \le \frac{1}{10}$ ، $\alpha \le \frac{1}{10}$

پس این تست ما برای $\bar{\beta}$ نیز برقرار است: $\forall p: \beta_p \le \frac{1}{10} \Rightarrow \bar{\beta} = E_{p\sim\pi}[\beta_p] \le \frac{1}{10}$

پس این تست برای فرض بیزی نیز برقرار است و از طرفی چون $\bar{\beta} \le \frac{1}{10}$ ، $\alpha \le \frac{1}{10}$ و از طرفی این خطاها از توزیع آزمون فرض دو توزیع P_0 و $\bar{P}$ می‌آیند داریم:

$$\frac{2}{10} \ge \alpha + \bar{\beta} = 1 - \left(E_{\bar{P}}[\varphi] - E_{P_0^{\otimes n}}[\varphi]\right) \ge 1 - TV(P_0^{\otimes n}, \bar{P})$$

$$\Rightarrow TV(P_0^{\otimes n}, \bar{P}) \ge \frac{8}{10} \quad \blacksquare$$

که دقت کنید بنا بر آخر استفاده مستقیم از لم ۱ در بخش (آ) بود که برای راحتی دوباره همان روند اثبات را طی نکردیم.

(د

$$\chi^2(P,Q) = \int \left(\frac{dP}{dQ}\right)^2 dQ - 1$$

در این بخش هم از $\bar{P}$ استفاده می‌کنیم، اینطور تعریف می‌کنیم:

$$\bar{P} = \mathbb{E}_{p\sim\pi}[P_p] \longrightarrow \bar{p}(x,y) = \mathbb{E}_{p\sim\pi}[p_p(x,y)]$$

داریم:

$$1 + \chi^2(\bar{P}, P_0) = \int \frac{\bar{P}(x,y)^2}{P_0(x,y)}\, dx\, dy$$

$$= \int \frac{\mathbb{E}_{p\sim\pi}(P_p(x,y))\, \mathbb{E}_{p'\sim\pi}[P_{p'}(x,y)]}{P_0(x,y)}\, dx\, dy$$

$$= \mathbb{E}_{p,p'\sim\pi} \int \frac{P_p(x,y)\, P_{p'}(x,y)}{P_0(x,y)}\, dx\, dy \qquad \blacksquare$$

حال برای حالت n متغیره تعریف می‌کنیم:

$$\bar{P}_n = \mathbb{E}_\pi[P_p^{\otimes n}] \longrightarrow \bar{P}_n(x^n, y^n) = \mathbb{E}_\pi\left[\prod_{i=1}^{n} P_p(x_i,y_i)\right]$$

داریم:

$$1 + \chi^2(\bar{P}_n, P_0^{\otimes n}) = \int \frac{\bar{P}_n(x^n,y^n)^2}{P_0^{\otimes n}(x^n,y^n)}\, dx^n\, dy^n$$

$$= \int \frac{\left(\mathbb{E}_{p\sim\pi}\left[\prod_{i=1}^{n} P_p(x_i,y_i)\right] \mathbb{E}_{p'\sim\pi}\left[\prod_{i=1}^{n} P_{p'}(x_i,y_i)\right]\right)}{\prod_{i=1}^{n} P_0(x_i,y_i)}\, dx^n\, dy^n$$

$$= \mathbb{E}_{p,p'\sim\pi} \int \prod_{i=1}^{n} \frac{P_p(x_i,y_i)\, P_{p'}(x_i,y_i)}{P_0(x_i,y_i)}\, dx^n\, dy^n$$

Subject : Date :

$$= \mathbb{E}_{P, P' \sim \pi^{\otimes 2}} \prod_{i=1}^{n} \int \frac{P_P(x_i, y_i)\, P_{P'}(x_i, y_i)}{P_0(x_i, y_i)}\, dx_i\, dy_i$$

با توجه به iid بودن $\rightarrow$

$$= \mathbb{E}_{P, P' \sim \pi^{\otimes 2}} \left[\left(\int \frac{P_P(x, y)\, P_{P'}(x, y)}{P_0(x, y)}\, dx\, dy \right)^n \right] \quad \blacksquare$$

(۵) دقت کنید که داریم:

$$P_\rho: \begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(0, \Sigma_\rho = \begin{bmatrix} 1 & \rho^T \\ \rho & I_d \end{bmatrix}\right) \longrightarrow P_\rho$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(0, \Sigma_{\rho'} = \begin{bmatrix} 1 & \rho'^T \\ \rho' & I_d \end{bmatrix}\right) \longrightarrow P_{\rho'}$$

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N(0, I_{d+1}) \longrightarrow P_0$$

مسئله ما در این بخش، محاسبه انتگرال روی چند متغیره گوسی است:

تعریف می‌کنیم $Z = (X, Y)$ پس $Z \in \mathbb{R}^{d+1}$ است، و فرض می‌کنیم از توزیع P_0 می‌آید، پس: $Z \sim N(0, I_{d+1})$

داریم:

$$\int \frac{P_\rho P_{\rho'}}{P_0} = \mathbb{E}_{Z \sim N(0, I_{d+1})} \left[\frac{P_\rho(z)}{P_0(z)} \cdot \frac{P_{\rho'}(z)}{P_0(z)} \right]$$

از طرفی داریم:

$$\frac{P_\rho(z)}{P_0(z)} = \det(\Sigma_\rho)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} z^T (\Sigma_\rho^{-1} - I) z\right) \longrightarrow \left(\text{same for } \frac{P_{\rho'}}{P_0}\right)$$

$$\Rightarrow \int \frac{P_\rho P_{\rho'}}{P_0} = \frac{1}{\sqrt{\det \Sigma_\rho \det \Sigma_{\rho'}}} \mathbb{E}_{Z \sim N(0, I)} \left[\exp\left(-\frac{1}{2} z^T (\Sigma_\rho^{-1} + \Sigma_{\rho'}^{-1} - 2I) z\right) \right]$$

از طرفی می‌دانیم که اگر $Z \sim N(0, I)$ و A ماتریسی متقارن با $A > -I$ باشد، داریم:

$$\mathbb{E}_{Z \sim N(0, I)} \left[-\frac{1}{2} z^T A z \right] = \frac{1}{\sqrt{\det(A + I)}} \longrightarrow (\text{با توجه به انتگرال گوسی})$$

پس در نهایت داریم:

$$\int \frac{P_p P_{p'}}{P_0} = \frac{1}{\sqrt{\det\Sigma_p \det\Sigma_{p'}}} \times \frac{1}{\det\left(\Sigma_p^{-1}+\Sigma_{p'}^{-1}-I\right)}$$

پس مسئله ما به محاسبه دترمینان چند ماتریس تقلیل یافته است:

$$\Sigma_p = \begin{bmatrix} 1 & p^T \\ p & I_d \end{bmatrix} \longrightarrow \det\Sigma_p = \det(I_d - pp^T) = 1-\|p\|^2$$

$$\Rightarrow \Sigma_p^{-1} = \begin{bmatrix} \frac{1}{1-\|p\|^2} & \frac{-p^T}{1-\|p\|^2} \\ -\frac{p}{1-\|p\|^2} & I_d + \frac{pp^T}{1-\|p\|^2} \end{bmatrix}$$

برای p' نیز همه اینها به طور مشابه تعریف و محاسبه می‌شوند.

با توجه به Σ_p^{-1} و $\Sigma_{p'}^{-1}$ برای ماتریس آخر داریم:

$$M = \Sigma_p^{-1} + \Sigma_{p'}^{-1} - I = \begin{bmatrix} a & -u^T \\ -u & S \end{bmatrix}$$

که در آن:

$$a = \frac{1}{1-\|p\|^2} + \frac{1}{1-\|p'\|^2} - 1 \qquad u = \frac{p}{1-\|p\|^2} + \frac{p'}{1-\|p'\|^2}$$

$$S = I_d + \frac{pp^T}{1-\|p\|^2} + \frac{p'p'^T}{1-\|p'\|^2}$$

با توجه به مطلب گفته شده در محاسبه det ماتریس های بلوکی داریم:

$$\det M = \det S\,(a - u^T S^{-1} u)$$

Subject : Date :

ابتدا $\det S$ را محاسبه می‌کنیم:

$$S = I + \underbrace{[p \;\; p']}_{U} \overbrace{\begin{bmatrix} \frac{1}{1-\|p\|^2} & 0 \\ 0 & \frac{1}{1-\|p'\|^2} \end{bmatrix}}^{C} \overbrace{\begin{bmatrix} p \\ p' \end{bmatrix}}^{U^T} = I + UCU^T$$

$$\Rightarrow \det S = \det(I + UCU^T) = \det(I_2 + CU^TU)$$

$$= \det \begin{bmatrix} \frac{1}{1-\|p\|^2} & \frac{p^Tp'}{1-\|p\|^2} \\ \frac{p^Tp'}{1-\|p'\|^2} & \frac{1}{1-\|p'\|^2} \end{bmatrix} = \frac{1-(p^Tp')^2}{(1-\|p\|^2)(1-\|p'\|^2)}$$

حال مقدار $a = u^TSu$ را محاسبه می‌کنیم:

(لم شرمن) $\longrightarrow$ $S^{-1} = I - U(\underbrace{C^{-1} + U^TU}_{K})^{-1}U^T$ $\longrightarrow$ داریم

از طرفی $\rightarrow (\underbrace{C^{-1} + U^TU}_{K})^{-1} = \underbrace{\begin{bmatrix} 1 & p^Tp' \\ p^Tp' & 1 \end{bmatrix}}_{K}^{-1} = \frac{1}{1-(p^Tp')^2} \begin{bmatrix} 1 & -p^Tp' \\ -p^Tp' & 1 \end{bmatrix}$

از طرفی u همان: $u = U \overbrace{\begin{bmatrix} \frac{1}{1-\|p\|^2} \\ \frac{1}{1-\|p'\|^2} \end{bmatrix}}^{w}$ است. پس:

$$S^{-1}u = (I - U(C^{-1}+U^TU)^{-1}U^T)Uw = U(w - (C^{-1}+U^TU)^{-1}U^TUw)$$

$$= U(w - K^{-1}U^TUw)$$

از طرفی $U^TUw = (K - C^{-1})w$ و همچنین $(K - C^{-1})w = Kw - 1$

Subject :

Date :

$$\text{پس} \rightarrow S^{-1}u = U(K^{-1}1)$$

$$\rightarrow u^T S^{-1} u = (U\omega)^T U(K^{-1}1) = \omega^T (U^T U) K^{-1} 1 = (\omega^T K - 1^T) K^{-1} 1$$

$$= \underbrace{\omega^T 1}_{} - \underbrace{1^T K^{-1} 1}_{}$$

$$= \frac{1}{1-\|p\|^2} + \frac{1}{1-\|p'\|^2} - \frac{2}{1+p^T p'}$$

پس در نهایت $\Rightarrow a - u^T S^{-1} u = \left(\frac{1}{1-\|p\|^2} + \frac{1}{1-\|p'\|^2} - 1\right) - \left(\frac{1}{1-\|p\|^2} + \frac{1}{1-\|p'\|^2} - \frac{2}{1+p^T p'}\right)$

$$\Rightarrow a - u^T S^{-1} u = \frac{2}{1+p^T p'} - 1 = \frac{1-p^T p'}{1+p^T p'}$$

پس به کل : $\det M = \det S\,(a - u^T S^{-1} u) = \frac{1-(p^T p)^2}{(1-\|p\|^2)(1-\|p'\|^2)} \times \frac{1-p^T p'}{1+p^T p'}$

$$\Rightarrow \det M = \frac{(1-p^T p')^2}{(1-\|p\|^2)(1-\|p'\|^2)}$$

پس در کل: $\int \frac{P_p P_{p'}}{P_0} = \frac{1}{\sqrt{(1-\|p\|^2)(1-\|p'\|^2)}} \times \frac{1}{\sqrt{\frac{(1-p^T p')^2}{(1-\|p\|^2)(1-\|p'\|^2)}}} = \frac{1}{1-p^T p'}$

پس حکم اثبات شد. ∎

MICRO

Subject : Date :

(و) طبق بخش ج همین سؤال، شرط لازم برای رسیدن به خطای حداکثر $\frac{1}{10}$ این است که: $\alpha, \beta_\sigma \le \frac{1}{10}$ و باید داشته باشیم:

$$TV(P_0^{\otimes n}, E_{\rho\sim\pi}[P_\rho^{\otimes n}]) \geq 0.8$$

در ادامه لم زیر را معرفی و اثبات می‌کنیم:

لم ۲: برای دو توزیع P و Q داریم: $TV(P,Q) \leq \frac{1}{2}\sqrt{\chi^2(Q\|P)}$

اثبات لم ۲: با توجه به تعریف TV و χ^2 به صورت f-divergence و نامساوی کوشی شوارتز داریم:

$$TV(Q,P) = \frac{1}{2}\int \left|\frac{dQ}{dP}-1\right| dP \leq \frac{1}{2}\left(\int \left(\frac{dQ}{dP}-1\right)^2 dP\right)^{\frac{1}{2}} = \frac{1}{2}\sqrt{\chi^2(Q\|P)}$$

پس لم اثبات شد. ■

حال با توجه به لم ۲ داریم:

$$0.8 \leq TV(P_0^{\otimes n}, E_\pi[P_\rho^{\otimes n}]) \leq \frac{1}{2}\sqrt{\chi^2(E_{\rho\sim\pi}[P_\rho^{\otimes n}], P_0^{\otimes n})}$$

طبق بخش (د)، برای عبارت آخر داریم:

$$\bar{P}_n := E_{\rho\sim\pi}[P_\rho^{\otimes n}]$$

$$1+\chi^2(\bar{P}_n, P_0^{\otimes n}) = E_{\rho,\rho'\sim\pi}\left[\left(\int \frac{P_\rho P_{\rho'}}{P_0}\right)^n\right] \overset{\text{طبق (ه)}}{=} E\left[(1-\rho^T\rho')^{-n}\right]$$

← حال طبق راهنمای سؤال عمل می‌کنیم و ρ و ρ' را به صورت یکنواخت از کره d بعدی rS^{d-1} با شعاع r انتخاب می‌کنیم.

پس داریم:

$$\rho = ru, \quad \rho' = rv \overset{iid}{\sim} Unif(S^{d-1})$$

MICRO

حال این فرض را جاگذاری می‌کنیم:

$$1+\chi^2(\bar{P}_n, P_0^{\otimes n}) = \mathbb{E}[(1-\rho^T\rho')^{-n}] = \mathbb{E}[(1-\tau^2 u^T v)^{-n}]$$

دقت کنید که u و v بردارهای یکه هستند. پس اگر تعریف کنیم: $Z := u^T v \in [-1,1]$

برای متغیر Z داریم: $\mathbb{E}[Z]=0$

$$\mathbb{E}[Z^2] = \mathbb{E}_u \mathbb{E}_v[(u^Tv)^2 | u] = \mathbb{E}_u[u^T \underbrace{\mathbb{E}[vv^T]}_{\substack{\text{تقارن} \\ \downarrow \\ \frac{1}{d} I_d}} u]$$

$$= \frac{1}{d}\mathbb{E}_u(\|u\|^2) = \frac{1}{d}$$

$$\Rightarrow Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \frac{1}{d}$$

داریم: $(1-\tau^2 u^T v)^{-n} = (1-\tau^2 Z)^{-n}$

از طرفی طبق تیلور داریم که اگر $|x| \le \frac{1}{2}$ داریم: $-\log(1-x) \le x + x^2$.

ما فرض می‌کنیم $|\tau^2 Z| \le \tau^2 \le \frac{1}{2}$. دقت کنید $\tau^2 \le \frac{1}{2}$ صرفا فرض و مسئله ما را tide ترمی کند و فرض بدی نیست.

پس داریم: $(1-\tau^2 Z)^{-n} \le \exp(n\tau^2 Z + n\tau^4 Z^2)$

Subject : Date :

پس داریم :

$$1+\chi^2(\bar{P}_n, P_0^{\otimes n}) \leq \mathbb{E}[\exp(n\sigma^2 Z + n\sigma^4 Z^2)] \leq e^{n\sigma^4}\, \mathbb{E}[e^{n\sigma^2 Z}]$$

از طرفی طبق MGF و بسط تیلور داریم :

$$\mathbb{E}[e^{\lambda Z}] = 1 + \sum_{k\geq 1} \frac{\lambda^{2k}}{(2k)!}\, \underbrace{\mathbb{E}[Z^{2k}]} \leq 1 + \sum_{k\geq 1} \frac{\lambda^{2k}}{(2k)!} \cdot \frac{(2k)!}{2^k k!\, d^k}$$

$$= 1 + \sum_{k\geq 1} \frac{1}{k!} \left(\frac{\lambda^2}{2d}\right)^k$$

داریم :

$$\mathbb{E}[Z^{2k}] \leq \frac{(2k)!}{2^k k!\, d^k}$$

$$= \exp\left(\frac{\lambda^2}{2d}\right)$$

با استفاده از روابط اثبات شده بالا داریم :

$$1+\chi^2(\bar{P}_n, P_0^{\otimes n}) \leq \exp\left(\frac{n^2\sigma^4}{2d} + n\sigma^4\right)$$

پس $\Rightarrow TV(P_0^{\otimes n}, \bar{P}_n) \leq \frac{1}{2}\sqrt{\chi^2(\bar{P}_n, P_0^{\otimes n})}$

$$\leq \frac{1}{2}\sqrt{\exp\left(\frac{n^2\sigma^4}{2d} + n\sigma^4\right) - 1}$$

حال دقت کنید برای اینکه شرایط مسئله برقرار باشد، باید $TV \geq 0.8$. حال شرایطی را در نظر بگیرید که $TV < 0.8$

Subject :

Date :

در این حالت، کافی است که: $\frac{1}{2}\sqrt{\exp\left(\frac{n^2\tau^4}{2d}+n\tau^4\right)-1} < 0.8$

$$\Leftrightarrow \exp\left(\frac{n^2\tau^4}{2d}+n\tau^4\right)-1 < 1.6$$

$$\Leftrightarrow \frac{n^2\tau^4}{2d}+n\tau^4 < \log(2.6) \approx 0.94$$

حال دقت کنید که اگر $n \leq \frac{c\sqrt{d}}{\tau^2}$ باشد، سمت چپ نامساوی برابر می‌شود با:

$$\frac{n^2\tau^4}{2d}+n\tau^4 \leqslant \frac{c^2 d\tau^4}{2d\tau^4}+\frac{c\sqrt{d}}{\tau^2}\tau^4 = \frac{c^2}{2}+c\sqrt{d}\,\tau^2$$

حال کافی است که عبارات، کافی کوچک بگیریم تا، $\frac{c^2}{2}+c\sqrt{d}\,\tau^4 < 0.94$ صدق کند.

پس برای $n \leq \frac{\sqrt{d}}{\tau^2}$ خواهیم داشت $TV < 0.8$ ← پس هیچ تستی به خطای حداکثر یک دهم نخواهد رسید.

پس حداقل تعداد نمونه مورد نیاز $n \geq \frac{c\sqrt{d}}{\tau^2}$ است. پس حداقل از مرتبه $\frac{d^\alpha}{\tau^\beta}$ خواهد بود که در آن $\alpha = \frac{1}{2}$ و $\beta = 2$ است. ■

(ز) همانطور که در قسمت اول بخش Problem Overview گفتیم، داریم که: $\begin{bmatrix}X\\Y\end{bmatrix} \sim \mathcal{N}\left(0, \Sigma_\rho = \begin{bmatrix}1 & \rho^T\\ \rho & I_d\end{bmatrix}\right)$

$$\longrightarrow Y|X=x \sim \mathcal{N}(\rho x, I_d - \rho\rho^T)$$

پس می‌توانیم از دید رگرسیون به Y نگاه کنیم و داریم:

$$Y = \rho X + \varepsilon \longrightarrow \varepsilon \sim \mathcal{N}(0, I_d - \rho\rho^T), \ \varepsilon \perp X$$

پس داریم $\Rightarrow$ $E_\rho[XY] = E[X(\rho X+\varepsilon)] = \rho E[X^2] + \underbrace{E[X\varepsilon]}_{0} = \rho \cdot 1 + 0 = \rho$

پس اثبات کردیم که $E[XY] = \rho$

پس منطقی است که تخمین گر $\hat\rho$ را به صورت مقابل تعریف کنیم: $\hat\rho = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$

دقت کنید که $\hat\rho$ تخمین گری unbiased است. با توجه به iid بودن متغیرها داریم:

$$Cov_\rho(\hat\rho) = \frac{1}{n} Cov_\rho(XY)$$

داریم: $Cov_\rho(XY) = E[XY(XY)^T] - \rho\rho^T = E[X^2 YY^T] - \rho\rho^T$

داریم $\longrightarrow$ $E[X^2 YY^T] = E[X^2(\rho X+\varepsilon)(\rho X+\varepsilon)^T]$

$$= E[X^4]\rho\rho^T + E[X^2]E[\varepsilon\varepsilon^T]$$

$$= 3\rho\rho^T + (I_d - \rho\rho^T) = I_d + 2\rho\rho^T$$

پس داریم:
$$\mathrm{Cov}_\mu(\hat{\mu}) = \frac{1}{n}\left(I_d + \mu\mu^T\right)$$

حال می‌توانیم $\mathbb{E}_\mu(\|\hat{\mu}-\mu\|^2)$ را با توجه به unbiased بودن $\hat{\mu}$ محاسبه کنیم:

$$\mathbb{E}_\mu[\|\hat{\mu}-\mu\|^2] = \operatorname{tr}\mathrm{Cov}_\mu(\hat{\mu}) = \frac{d+\|\mu\|^2}{n} \leq \frac{d+1}{n} = \Theta\left(\frac{d}{n}\right)$$

از طرفی تست خود را با توجه به این تخمین‌گر تعریف می‌کنیم:
$$\varphi = \mathbb{1}\left\{\|\hat{\mu}\| \geq \frac{r}{2}\right\}$$

حال خطاها را حساب می‌کنیم:

$$\mathrm{I} \to P_0(\varphi=1) = P_0\left(\|\hat{\mu}\|^2 \geq \frac{r^2}{4}\right) \leq \frac{\mathbb{E}_0[\|\hat{\mu}\|^2]}{\frac{r^2}{4}} = \frac{4d}{nr^2} \to \alpha$$

$$\mathrm{II} \to P_\mu(\varphi=0) \leq P_\mu\left(\|\hat{\mu}-\mu\| \geq \frac{r}{2}\right) \leq \frac{\mathbb{E}_\mu[\|\hat{\mu}-\mu\|^2]}{\frac{r^2}{4}} = \frac{4(d+\|\mu\|^2)}{nr^2} \leq \frac{4(d+1)}{nr^2} \to \beta$$

حال برای اینکه هر دو خطای α و β کمتر از $\frac{1}{10}$ باشند، باید داشته باشیم:

$$\frac{4(d+1)}{nr^2} \leq \frac{1}{10} \longrightarrow n \geq \frac{40(d+1)}{r^2} \to n \sim \frac{Cd}{r^2}$$

پس تعداد نمونه مورد نیاز ما در مرتبه $\frac{d}{r^2}$ خواهد بود. که همانطور که صورت مسئله گفته با کران پایین مطابقت ندارد. چون در قسمت قبل به کران $\frac{\sqrt{d}}{r^2}$ رسیده بودیم! $\blacksquare$

ج) دقت کنید که Z_1 و Z_2 دقیقا مانند $\hat{\mu}$ در قسمت قبل تعریف شده‌اند، تنها با $\frac{1}{2}$ تعداد نمونه‌های آن‌ها.

پس به طور مشابه قسمت قبل می‌توان گفت که داریم:

$$E_\mu[Z_1] = E_\mu[Z_2] = \mu$$

$$Cov(Z_1) = Cov(Z_2) = \frac{1}{\frac{n}{2}}(I_d + \mu\mu^T) = \frac{2}{n}(I_d + \mu\mu^T)$$

حال برای $T = Z_1^T Z_2$ داریم:

$$E_\mu[T] = E_\mu[Z_1^T Z_2] = E_\mu[Z_1]^T E_\mu[Z_2] = \mu^T\mu = \|\mu\|^2$$

حال با توجه به unbiased بودن Z_1 و Z_2، تعریف می‌کنیم:

$$Z_1 = \mu + \varepsilon_1 \quad , \quad Z_2 = \mu + \varepsilon_2$$

که ε_1 و ε_2 مستقل از هم با میانگین 0 و ماتریس کواریانس $\frac{2}{n}(I_d + \mu\mu^T)$ هستند:

برای واریانس T داریم:

$$T = \|\mu\|^2 + \mu^T(\varepsilon_1 + \varepsilon_2) + \varepsilon_1^T \varepsilon_2$$

$$\rightarrow \quad Var_\mu(T) = Var(\mu^T\varepsilon_1) + Var(\mu^T\varepsilon_2) + Var(\varepsilon_1^T\varepsilon_2) = 2\,Var(\mu^T\varepsilon_1) + Var(\varepsilon_1^T\varepsilon_2)$$

از طرفی می‌دانیم:

$$Var(\mu^T\varepsilon_1) = \mu^T\Sigma\mu = \frac{2}{n}(\|\mu\|^2 + \|\mu\|^4)$$

و برای دو بردار مستقل با میانگین 0 و کواریانس ماتریس Σ می‌دانیم که:

$$Var(\varepsilon_1^T\varepsilon_2) = tr(\Sigma^2) =$$

$$\Rightarrow \quad Var_\mu(T) = \frac{4}{n}(\|\mu\|^2 + \|\mu\|^4) + tr\left(\left[\frac{2}{n}(I + \mu\mu^T)\right]^2\right) =$$

$$= \boxed{\frac{4}{n}(\|\mu\|^2 + \|\mu\|^4) + \frac{4}{n^2}(d + 2\|\mu\|^2 + \|\mu\|^4)}$$

دقت کنید که اگر $\mu = 0$ باشد، $Var_0(T) = \frac{4d}{n^2}$ خواهد بود.

حال تست را از روی T به این صورت تعریف می‌کنیم: $\text{Reject } H_0 \text{ if } T \geq \frac{\varepsilon^2}{2}$

حال با توجه به این تست به بررسی خطاها می‌پردازیم:

$$\alpha \to P_0\left(T \geq \frac{\varepsilon^2}{2}\right) \leq \frac{Var_0(T)}{\left(\frac{\varepsilon^2}{2}\right)^2} = \frac{\frac{4d}{n^2}}{\frac{\varepsilon^4}{4}} = \frac{16d}{n^2\varepsilon^4}$$

$$\beta \to P_p\left(T < \frac{\varepsilon^2}{2}\right) = P_p\left(\left|T - \|p\|^2\right| \geq \|p\|^2 - \frac{\varepsilon^2}{2}\right) \leq \frac{Var_p(T)}{\left(\|p\|^2 - \frac{\varepsilon^2}{2}\right)^2} \leq \frac{Var_p(T)}{\left(\frac{\varepsilon^2}{2}\right)^2}$$

حال برای کران بالای $Var_p(T)$ دقت کنید که $\|p\|^2 \leq 1$ پس طبق فرمول صفحه قبل و اعمال این کران خواهیم داشت:

$$Var_p(T) \leq \frac{8}{n} + \frac{4(d+3)}{n^2}$$

$$\text{پس} \to \beta \leq \frac{32}{n\varepsilon^4} + \frac{16(d+3)}{n^2\varepsilon^4}$$

حال دقت کنید اگر تعداد نمونه‌ها در مرتبه $\frac{\sqrt{d}}{\varepsilon^2}$ باشد، مثلاً قرار دهیم: $n = k\frac{\sqrt{d}}{\varepsilon^2}$ خواهیم داشت:

$$\alpha \leq \frac{16}{k^2} \quad , \quad \beta \leq \frac{32}{k\sqrt{d}\varepsilon^2} + \frac{16}{k^2}\left(1+\frac{3}{d}\right) \leq \frac{32}{k\sqrt{d}\varepsilon^2} + \frac{32}{k^2}$$

کافی است که k از عدد ثابتی بیشتر باشد تا هر دو خطا کمتر از $\frac{1}{10}$ شوند: $\boxed{n \geq C\frac{\sqrt{d}}{\varepsilon^2}}$

که این تعداد نمونه با کران پایین همخوانی دارد. ∎

Subject :   Date :

ط) تعریف می‌کنیم $S = \text{sign}(X)$ پس $SX = |X|$ حال دقت کنید برای $\mathbb{E}_\rho$ داریم:

$$\mathbb{E}_\rho[SY] = \mathbb{E}_\rho[S(\rho X + \varepsilon)] = \rho \underbrace{\mathbb{E}_\rho[SX]}_{\sqrt{\frac{2}{\pi}}} + \underbrace{\mathbb{E}_\rho[S\varepsilon]}_{0} = \rho\sqrt{\frac{2}{\pi}}$$

برای راحتی قرار می‌دهیم $c = \sqrt{\frac{2}{\pi}}$ پس $\mathbb{E}_\rho[SY] = c\rho$

حال با تعریف مشابه مانند بخش قبل برای Z_1 و Z_2 داریم:

$$Z_1 = \frac{2}{n}\sum_{i=1}^{\frac{n}{2}} \tilde{X}_i Y_i \quad , \quad Z_2 = \frac{2}{n}\sum_{i=\frac{n}{2}+1}^{n} \tilde{X}_i Y_i$$

$$\Rightarrow \mathbb{E}_\rho[Z_1] = \mathbb{E}_\rho[Z_2] = c\rho$$

برای محاسبه Cov آن‌ها دقت کنید که: $SY = \rho|X| + S\varepsilon$

$$\rightarrow \mathbb{E}[(SY)(SY)^T] = \rho\rho^T\mathbb{E}[|X|^2] + \mathbb{E}[S\varepsilon\varepsilon^T S] = \rho\rho^T + (I_d - \rho\rho^T) = I_d$$

$$\Rightarrow Cov(SY) = \mathbb{E}[(SY)(SY)^T] - \mathbb{E}[SY]\,\mathbb{E}[SY]^T = I_d - c^2\rho\rho^T$$

پس با توجه اینکه Z_1 و Z_2 جمع کپی‌های SY است داریم:

$$Cov(Z_1) = Cov(Z_2) = \frac{2}{n}\left(I_d - c^2\rho\rho^T\right)$$

حال با تعریف T مانند قبل؛ $T_{sign} := Z_1^T Z_2$ داریم: $\mathbb{E}_\rho[T_{sign}] = \mathbb{E}[Z_1]^T\,\mathbb{E}[Z_2] = c^2\|\rho\|^2$

Subject :   Date :

مانند قبل تعریف می کنیم:

$$Z_1 = c\rho + \varepsilon_1 \quad , \quad Z_2 = c\rho + \varepsilon_2$$

که مانند قبل: $\mathrm{Cov}(\varepsilon_1) = \mathrm{Cov}(\varepsilon_2) = \underbrace{\frac{2}{n}\left(I_d - c^2\rho\rho^T\right)}_{\Sigma}$, $\mathbb{E}[\varepsilon_1] = \mathbb{E}[\varepsilon_2] = 0$

$$\longrightarrow \mathrm{Var}_\rho(T_{sign}) = 2c^2\rho^T\Sigma\rho + \mathrm{Var}(\varepsilon_1^T\varepsilon_2)$$

$$= 2c^2\rho^T\Sigma\rho + \mathrm{tr}(\Sigma^2)$$

$$= \frac{2}{n}\left(\|\rho\|^2 - c^2\|\rho\|^4\right) + \frac{4}{n^2}\left(d - 2c^2\|\rho\|^2 + c^4\|\rho\|^4\right)$$

و همچنان اگر $\rho = 0$ داریم: $\mathrm{Var}_\rho(T_{sign}) = \frac{4d}{n^2}$

پس برای خطاها مانند قبل داریم:

$$\alpha \leq \frac{16d}{c^4 n^2 \tau^4} \qquad \beta \leq \frac{16}{c^2 n \tau^4} + \frac{16(d+1)}{c^4 n^2 \tau^4}$$

که مانند قبل می توان نشان داد برای $\alpha, \beta \leq 0.1$ کافی است $n \geq C\frac{\sqrt{d}}{\tau^2}$. مانند حالت قبل!

سه کاربردهای این مطلب:

طبق کتنی 24.1 کتاب Polyanskiy ، کوانتیزه کردن بیت های گاوسی و برای reconstruct کردن داده ها. به طور کلی این علامت X

با اینکه اشاره ندارد همچنان اطلاعاتی نسبت به ρ در Y را دارد و جهت می دهد که برای کاربردهای مانند Low-rate sensing

MICRO

communication-constrained testing استفاده می‌شود

همچنین برای privacy زمانی که تنها علامت را در دسترس داریم و مکانیزم انداریم به علامت، همچنان فاصله‌ی حملی رگرسیون را برقرار کرده و

ما با تعداد موردنیاز $\frac{d}{\epsilon^2}$ داده می‌توانیم تست خود را با خطای کم بازیابی کنیم

همچنین همانطور که در بین حل گفته شد از بخش 7.5 کتاب نیز برای اثبات بخشی از استدلال استفاده شد.