« بسمه تعالی »

سوالات را به دقت خوانده و به تمامی اجزای آن پاسخ دهید.

۱. **نامساوی.**

(آ) برای توزیع‌های ضربی $P(x_1,\cdots,x_n)=P_1(x_1)\times\cdots\times P_n(x_n)$ و $Q(x_1,\cdots,x_n)=Q_1(x_1)\times\cdots\times Q_n(x_n)$ رابطه زیر را نشان دهید:

$$\mathsf{TV}\left(P,Q\right)\le\sum_{i=1}^{n}\mathsf{TV}\left(P_i,Q_i\right)$$

(ب) رابطه زیر را بین TVو واگرایی χ^2 نشان دهید:

$$\mathsf{TV}(P,Q)\le c\sqrt{\frac{\chi^2(P,Q)}{1+\chi^2(P,Q)}}$$

که در آن $c>0$ عدد ثابتی است که میبایست کوچکترین مقدار آن را که نامساوی فوق برقرار باشد، را بدست آورید. اثبات نامساوی فوق برای یک c دلخواه بدون اثبات بهینه‌بودن آن حائز ٪۹۰ نمره این بخش است.

(ج) **TVبین دو توزیع برداری گوسی** فرض کنید C یک ماتریس مثبت معین با مقادیر ویژه $c_1,\cdots,c_d$ باشد. نشان دهید:

$$\mathsf{TV}\left(\mathcal{N}(\mathbf{0},C),\mathcal{N}(\mathbf{0},I_d)\right)\le\sqrt{2\sum_{i=1}^{d}(\ln c_i)^2}$$

راهنمایی: تجزیه قطری $C=\sum_{i=1}^{d}c_iu_iu_i^\top$ را در نظر بگیرید. به ازای این تجزیه قطری توان ماتریس C به صورت زیر مشخص می‌شود:

$$C^t=\sum_{i=1}^{d}c_i^tu_iu_i^\top$$

در این صورت با تغییر t بین صفر تا یک، یک *پیمایش* بین ماتریسهای همانی و C در فضای ماتریسها خواهیم داشت. برای ماتریس مثبت معین دلخواه B فرض کنید که $f_B(x)$ تابع چگالی احتمال بردار گوسی تصادفی با میانگین صفر و ماتریس کوواریانس B باشد. ابتدا نشان دهید

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

و سپس سعی کنید که از این رابطه در تعریف TV به نحو مناسبی استفاده نمایید.

۲. **کاربرد فرم وردشی در ابعاد بالا** در این مسئله به کاربرد فرم وردشی در یکی از مسائل مربوط به درس احتمالات ابعاد بالا HDP می‌پردازیم.

(آ) فرم وردشی زیر (موسوم به گیبس) برای تابع مولد گشتاور که ارتباط تنگاتنگی با فرم Donsker-Varadhan دارد را اثبات نمایید:

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

(ب) فرض کنید X یک متغیر تصادفی دلخواه روی دامنه $\mathcal{X}$ باشد. همچنین Θ زیرمجموعه‌ای از فضای $\mathbb{R}^d$ باشد. برای مثال $\mathcal{X}$ می‌تواند فضای ویژگی‌ها و Θ فضای پارامترها باشد. همچنین فرض کنید که تابع $f:\mathcal{X}\times\Theta\mapsto\mathbb{R}$ دارای این خاصیت باشد که $\mathbb{E}_X\exp(f(X,\theta)$ برای هر θ متناهی باشد. فرض کنید μ یک توزیع احتمال ثابت روی فضای پارامتر باشد. نشان دهید

$$\mathbb{P}_X\left[\exists\nu:\mathbb{E}_\nu[f(X,\theta]\ge\mathbb{E}_\nu\left[\log(\mathbb{E}_X\exp(f(X,\theta)))\right]+D(\nu\|\mu)+t\right]\le\exp(-t)$$

به عبارت دیگر احتمال اینکه توزیع دلخواه ν روی فضای پارامتر وجود داشته باشد به طوریکه نامساوی داده‌شده برقرار باشد، حداکثر e^{-t} است.

(ج) از اینجا به کاربرد در HDP می‌پردازیم. فرض کنید $\mathbf{X}$ یک ماتریس تصادفی $m\times n$ باشد که درایه‌های آن همگی گوسی استاندارد و مستقل از هم هستند. هدف از این قسمت بررسی نرم اپراتوری ماتریس $\mathbf{X}$ است که به صورت زیر قابل بیان است:

$$\|\mathbf{X}\|=\sup_{\mathbf{u}\in\mathbb{S}^{m-1},\mathbf{v}\in\mathbb{S}^{n-1}}\mathbf{u}^\top\mathbf{X}\mathbf{v}$$

در اینجا منظور از $\mathbb{S}^{d-1}$ مجموعه بردارهای dـ بعدی به طول یک است . هدف نشان‌دادن رابطه زیر است:

$$\mathbb{P}\left[\|\mathbf{X}\| \le c\sqrt{m+n+t}\right] \ge 1-\exp(-t) \quad (۱)$$

است که در آن c عدد ثابتی است که در طول محاسبه بدست خواهید آورد. بدین منظور قرار دهید

$$\Theta = (1+\epsilon)\mathbb{B}^m \times (1+\epsilon)\mathbb{B}^n$$

که در آن منظور از $\mathbb{B}^d$ کره واحد dـ بعدی است. همچنین فرض کنید μ حاصلضرب دو توزیع احتمال یکنواخت روی $(1+\epsilon)\mathbb{B}^m$ و $(1+\epsilon)\mathbb{B}^n$ باشد. همچنین برای هر زوج بردار به طول واحد $(\mathbf{u},\mathbf{v}) : \mathbf{u} \in \mathbb{S}^{m-1}, \mathbf{v} \in \mathbb{S}^{n-1}$ توزیع احتمال $\nu_{\mathbf{u},\mathbf{v}}$ نظیر می‌کنیم که برابر است با حاصلضرب دو توزیع احتمال یکنواخت روی کره‌های $\{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}-\mathbf{u}\| \le \epsilon\}$ و $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}-\mathbf{v}\| \le \epsilon\}$. در قدم بعد ابتدا $D(\nu_{\mathbf{u},\mathbf{v}}\|\mu)$ را محاسبه کرده و سپس با تعریف مناسب تابع $f(\mathbf{X};(\mathbf{u},\mathbf{v}))$ در قسمت قبل،کمی محاسبات ساده با متغیر گوسی و در نهایت مقداردهی مناسب ϵ ، رابطه (۱) را بدست آورید.

۳. **تست استقلال و آزمون فرض مرکب** . در این مسئله هدف بررسی مستقل بودن دو متغیر تصادفی مشترکا گوسی از طریق دسترسی به n نمونه از این زوج است. بدین منظور مسئله را در ساختار زیر بررسی می‌کنیم. فرض کنید X یک متغیر اسکالر گوسی استاندارد باشد و $\mathbf{Y}$ یک بردار گوسی dـ بعدی باشد به طوریکه با X مشترکا گوسی باشد و به شرط $X = x$ یک بردار گوسی با توزیع $\mathbb{P}_\rho = \mathcal{N}\left(\rho x, I_d - \rho\rho^\top\right)$ باشد که در آن ρ بردار ضرایب همبستگی بردار $\mathbf{Y}$ با X است. ملاحظه کنید که $\mathbf{Y}$ به تنهایی یک بردار گوسی استاندارد است (چرا؟). هدف از تست استقلال اینست که بررسی کنیم که آیا بردار همبستگی صفر است یا خیر و برای جواب‌دادن به این سوال به چند نمونه از توزیع $\mathbb{P}_\rho(x,\mathbf{y})$ نیاز است. تعریف دقیق‌تر در ادامه می‌آید.

(آ) ابتدا حالت $d=1$ را در نظر بگیرید و آزمون فرض دوتایی زیر را در نظر بگیرید:

$$H_0 : \rho = 0, \quad \text{vs} \quad H_1 : \rho = \tau$$

فرض کنید که به n زوج نمونه $\{(X_1,Y_1),\cdots,(X_n,Y_n)\}$ از توزیع $\mathbb{P}_\rho$ دسترسی داریم. هدف بررسی تعداد نمونه لازم بر حسب τ است به طوری‌که خطای نوع یک و دو هر دو کمتر از یک‌دهم باشند یعنی:

$$\mathbb{P}_0[\mathsf{accept}H_1] \le \frac{1}{10}, \quad \mathbb{P}_\tau[\mathsf{accept}H_0] \le \frac{1}{10}$$

نشان دهید که تعداد نمونه مورد نیاز حداقل از مرتبه $\frac{1}{\tau^2}$ است. ضمنا یک تست (حتی‌المقدور ساده) ارایه کنید که با این تعداد نمونه به خطای مطلوب حداکثر یک‌دهم برسد.

(ب) حال آزمون فرض دوتایی زیر را در بعد دلخواه d در نظر بگیرید:

$$H_0 : \rho = [0,\cdots,0], \quad \text{vs} \quad H_1 : \rho = [\tau,0,\cdots,0]^\top$$

بدون نوشتن محاسبات پیچیده استدلال کنید که جواب مسئله فوق با حالت یک‌بعدی یکسان است. یعنی تعداد نمونه‌های مورد نیاز از مرتبه $\frac{1}{\tau^2}$ خواهد بود.

(ج) **آزمون فرض مرکب** حال آزمون فرض دوتایی زیر را در نظر بگیرید:

$$H_0 : \rho = \mathbf{0}, \quad \text{vs} \quad H_1 : \rho \in \mathbb{B}_\tau^c := \{\rho : \|\rho\| \ge \tau\}$$

دقت کنید که در فرض H_1 توزیع داده‌ها ثابت نیست و هر توزیعی با اندازه بردار همبستگی بیشتر از حد آستانه τ درون این فرض قرار دارد. در این حالت گوییم که یک تست دارای احتمای خطای نوع یک و دو کمتر از یک‌دهم است، اگر

$$\mathbb{P}_0[\mathsf{accept}H_1] \le \frac{1}{10}, \quad \mathbb{P}_\rho[\mathsf{accept}H_0] \le \frac{1}{10}, \;\; \forall \rho \in \mathbb{B}_\tau^c.$$

بدست آوردن حداقل تعداد نمونه مورد نیاز برای تمیزدادن دو فرضیه H_0 و H_1 به سادگی حالت آزمون فرض ساده نیست. یک راه مرسوم تحلیل مسئله فوق از طریق ارتباط‌دادن آن با یک آزمون فرض دوتایی بیزی است. بدین منظور فرض کنید π یک توزیع احتمال دلخواه با دامنه فضای فرضیه H_1 یعنی $\mathbb{B}_\tau^c$ باشد. حال آزمون فرض دوتایی جدید بیزی را به شکل زیر می‌سازیم. ابتدا یکی از دو فرض را با احتمال یکسان انتخاب می‌کنیم. در فرض صفر، همچنان n نمونه مستقل از توزیع $\mathbb{P}_0$ دریافت می‌کنیم. در فرض جایگزین، ابتدا یک ρ با احتمال $\pi(\rho)$ از $\mathbb{B}_\tau^c$ انتخاب می‌کنیم و سپس n نمونه مستقل از توزیع $\mathbb{P}_\rho$ دریافت می‌کنیم. فرض کنید که تست موفقی برای آزمون فرض مرکب با خطای کمتر از یک‌دهم در اختیار ماست. نشان دهید که همان تست دارای خطای کمتر از یک‌دهم برای آزمون فرض بیزی نیز هست. سپس نشان دهید

$$\mathsf{TV}(\mathbb{P}_0^{\otimes n}, \mathbb{E}_{\rho\sim\pi}[\mathbb{P}_\rho^{\otimes n}]) \ge 0.8.$$

دقت کنید که در اینجا منظور از $\mathbb{E}_{\rho\sim\pi}[\mathbb{P}_\rho^{\otimes n}]$ توزیع احتمال حاشیه‌ای القاشده روی n نمونه پس از انتخاب تصادفی ρ است و قابل بیان به شکل زیر نیز هست:

$$\mathbb{E}_{\rho\sim\pi}[\mathbb{P}_\rho^{\otimes n}](x^n,\mathbf{y}^n) = \int \left(\prod_i \mathbb{P}_\rho(x_i,\mathbf{y}_i)\right)\pi(\rho)d\rho$$

(د) محاسبه و حتی کران‌زدن فاصله TV بین توزیع‌های قسمت قبل به سادگی محاسبه TV در حالت ضربی نیست. شناخته‌شده‌ترین راه استفاده از واگرایی χ^2 و ارتباط آن با TV است. متاسفانه واگرایی‌های دیگر مانند KL یا هلینجر در اینجا مفید نیستند.
نشان دهید:

$$1 + \chi^2(\mathbb{E}_\pi[\mathbb{P}_\rho], \mathbb{P}_0) = \mathbb{E}_{(\rho,\rho')\sim\pi^{\otimes 2}} \left[\int \frac{\mathbb{P}_\rho(x,\mathbf{y})\mathbb{P}_{\rho'}(x,\mathbf{y})}{\mathbb{P}_0(x,\mathbf{y})} dx d\mathbf{y} \right]$$

که در آن ρ, ρ' مستقل از یکدیگر و از روی توزیع π تولید شده‌اند. سپس رابطه زیر را نتیجه بگیرید:

$$1 + \chi^2(\mathbb{E}_\pi[\mathbb{P}_\rho^{\otimes n}], \mathbb{P}_0^{\otimes n}) = \mathbb{E}_{(\rho,\rho')\sim\pi^{\otimes 2}} \left[\left(\int \frac{\mathbb{P}_\rho(x,\mathbf{y})\mathbb{P}_{\rho'}(x,\mathbf{y})}{\mathbb{P}_0(x,\mathbf{y})} dx d\mathbf{y} \right)^n \right]$$

(ه) نشان دهید:

$$\int \frac{\mathbb{P}_\rho(x,\mathbf{y})\mathbb{P}_{\rho'}(x,\mathbf{y})}{\mathbb{P}_0(x,\mathbf{y})} dx d\mathbf{y} = \frac{1}{1 - \rho^\top \rho'}$$

حل این قسمت صرفا محاسباتی است. در صورتیکه قادر به اثبات این قسمت نبودید، از آن صرف نظر کرده و صرفا در قسمت‌های بعدی از نتیجه آن استفاده نمایید.

(و) با استفاده از قسمت‌های قبل نشان دهید که تعداد نمونه مورد نیاز برای رسیدن به خطای حداکثر یک‌دهم حداقل از مرتبه $\frac{d^\alpha}{\tau^\beta}$ است که در آن α, β اعدادی مثبت هستند که می‌بایست مقادیر آن‌ها را بدست بیاورید.
راهنمایی: توزیع π را یک توزیع یکنواخت روی سطح کره dـ بعدی $\tau\mathbb{S}^{d-1}$ به شعاع τ بگیرید. در صورت لزوم می‌توانید از بسط تیلور برای تقریب‌زدن استفاده نمایید.

(ز) در ادامه به بررسی چفت (tight) بودن کران پایین بدست آمده در قسمت قبل می‌پردازیم. بدین منظور دو تست مختلف معرفی می‌کنیم. یک راه اولیه اینست که سعی کنیم ابتدا بردار ρ را تخمین بزنیم و سپس با محاسبه اندازه تخمین آن در مورد قبول یا رد هر فرضیه تصمیم بگیریم. ابتدا نشان دهید $\mathbb{E}_\rho[X\mathbf{Y}] = \rho$. با استفاده از این نکته یک تخمین‌گر تجربی $\hat{\rho}$ معرفی کنید و نشان دهید که مجذور خطای آن (یعنی $\mathbb{E}[(\rho - \hat{\rho})^2]$) از مرتبه $\frac{d}{n}$ است. از اینجا نتیجه بگیرید که تعداد نمونه مورد نیاز این روش از مرتبه $\frac{d}{\tau^2}$ خواهد بود که با کران پایین مطابقت ندارد.

(ح) روش تخمین و تست در قسمت قبل دارای این ایراد است که در مسئله تست فقط تصمیم‌گیری در مورد اندازه بردار همبستگی مهم است و نه مقدار و جهت آن. با توجه به اینکه جهت بردار همبستگی در طول زمان ثابت است، روش زیر پیشنهاد می‌شود. ابتدا داده‌ها را به دو قسمت مساوی با اندازه $\frac{n}{2}$ تقسیم می‌کنیم. قرار دهید

$$Z_1 = \frac{2}{n} \sum_{i=1}^{n/2} X_i \mathbf{Y}_i, \quad Z_2 = \frac{2}{n} \sum_{i=n/2+1}^{n} X_i \mathbf{Y}_i$$

نشان دهید

$$\mathbb{E}_\rho[Z_i] = \rho$$

قرار دهید $T = Z_1^\top Z_2$. نشان دهید $\mathbb{E}_\rho[T] = \|\rho\|^2$ با استفاده از T یک تست ارایه کنید و تحلیل کنید. در تحلیلتان می‌توانید به نامساوی ساده‌ای مانند چبیشف اکتفا کنید. مشاهده کنید که تعداد نمونه‌های لازم با کران پایین همخوانی دارد.

(ط) حال فرض کنید که ما به جای دسترسی به مقادیر واقعی X_i ها فقط به علامت انها دسترسی داریم. یعنی مشاهده ما شامل n نمونه $\{(\widetilde{X}_1, \mathbf{Y}_1), \cdots, (\widetilde{X}_n, \mathbf{Y}_n)\}$ است که در آن $\widetilde{X}_i = \mathrm{sign}(X_i)$. به نظرتان این مطلب در کجا قابل کاربرد است (میتوانید کتاب پولیانسکی را ورق بزنید!).
نشان دهید که استفاده از تست قسمت قبل که در آن X_iها با علامتشان جایگزین شده‌اند، همچنان می‌تواند به کران پایین برسد.