باسمه تعالی

مباحثی در آمار

سری اول (استنباط فراوانی‌گرا و بیزی)

---

۱. **لم نیمن‌ــ پیرسون**

آزمون فرض یک روش آماری است برای تصمیم‌گیری راجع به رد کردن یا نکردن یک فرض صفر (H_0) در برابر یک فرض بدیل (H_1) ، بر اساس داده مشاهده شده. منظور از خطای نوع یک رد شدن فرض صفر در صورت درست بودن آن است و منظور از خطای نوع دو رد نشدن فرض صفر در صورت غلط بودن این فرض است.

$$\alpha = \mathbb{P}(H_0 \text{ رد شدن} | H_0 \text{ درستی})$$
$$\beta = \mathbb{P}(H_0 \text{ رد نشدن} | H_0 \text{ غلط بودن})$$

به α سطح معنی‌داری آزمون و به $1-\beta$ توان آزمون می‌گویند.

فرض کنید داده x مشاهده شده و فرض صفر و بدیل ما راجع به تابع چگالی توزیع x است:

$$\begin{cases} H_0: & x \sim P_0 \\ H_1: & x \sim P_1 \end{cases}$$

الف) لم نیمن‌ــ پیرسون را ثابت کنید:

عددی مانند t وجود دارد که در بین آزمون‌های با سطح معنی‌داری α ، آزمون زیر بیشترین توان را دارد:

$$\frac{P_1(x)}{P_0(x)} > t \implies H_0 \text{ رد شدن}$$

ب) این لم را برای حالت زیر اعمال کنید و نتیجه را توصیف کنید:

$$\begin{cases} H_0: & x_1, \cdots, x_n \sim \mathcal{N}(0,1) \\ H_1: & x_1, \cdots, x_n \sim \mathcal{N}(\mu,1) \quad (\mu > 0) \end{cases}$$

۲. **آزمون فرض بیزی**

الف) فرض کنید $x \sim \mathcal{N}(\mu, \sigma^2)$ و برای μ توزیع پیشینی در نظر می‌گیریم که احتمال صفر شدن آن مثبت باشد: $\mu \sim p\delta_0 + (1-p)\mathcal{N}(0, s^2)$. در این‌جا δ_0 اندازه اتمی دلتای دیراک در نقطه صفر است. توزیع پسین $g(\mu|x)$ را محاسبه کنید.

ب) توضیح دهید که چگونه می‌توان یک آزمون فرض بیزی برای بررسی فرض $\mu = 0$ در مقابل فرض $\mu \neq 0$ طراحی کرد.

۳. **مشاهدات بیشتر و افزایش اطمینان**
الف) فرض کنید $x \sim \mathcal{N}(\mu, \sigma^2)$ و $\mu \sim \mathcal{N}(0, s^2)$. توزیع پسین $g(\mu|x)$ را محاسبه کنید و نشان دهید واریانس توزیع پسین کمتر از واریانس توزیع پیشین است.
ب) آیا می‌توانید مثالی ارائه دهید که واریانس توزیع پسین از توزیع پیشین بیشتر باشد؟
ج) نشان دهید به طور متوسط مشاهده داده عدم اطمینان ما را کاهش می‌دهد! به این معنی که اگر واریانس توزیع پسین را متغیری تصادفی در نظر بگیریم (به خاطر تصادف موجود در داده)، امید ریاضی آن کوچک‌تر یا مساوی واریانس توزیع پیشین است:

$$\mathbb{E}_x[\mathrm{Var}[\theta|x]] \leq \mathrm{Var}[\theta]$$

۴. **آماره بسنده**
در مدل پارامتری $f_\theta(x)$ آماره $s = s(x)$ بسنده است.
الف) نشان دهید آماردان فراوانی‌گرا برای استنباط با روش بیشترین درست‌نمایی فقط به مقدار s نیاز دارد.
ب) نشان دهید آماردان بیزگرا هم برای محاسبه توزیع پسین نیازی به کل داده x ندارد و دانستن s برای او کافی است.

۵. **کمیت لولا**
منظور از کمیت لولا تابعی از داده‌ها و پارامترهای مدل است که توزیع آن به پارامترها بستگی نداشته باشد.
فرض کنید $Y_1, \cdots, Y_n$ نمونه‌ای $i.i.d.$ از تابع چگالی $f_\theta(x) = \frac{1}{\theta} e^{\frac{-x}{\theta}}$ باشد. اگر $\hat{\theta}$ برآوردگر بیشترین درست‌نمایی باشد، ابتدا توضیح دهید چگونه می‌توان با توجه به توزیع مجانبی $\hat{\theta}$ یک بازه اطمینان تقریبی برای θ ارائه داد.
سپس نشان دهید $\frac{\hat{\theta}}{\theta}$ یک کمیت لولاست، توزیع آن را به دست آورید و توضیح دهید چگونه می‌توان به کمک آن یک بازه اطمینان دقیق ساخت.

۶. **توزیع پیشین جفریز**
توزیع پیشین جفریز یک توزیع ناآگاهنده روی فضای پارامتر است که متناسب است با ریشه دوم اطلاع فیشر: $g^{\mathrm{Jeff}}(\theta) \propto \sqrt{I_\theta}$.
الف) نشان دهید توزیع پیشین جفریز تحت تغییر پرمایش خوش رفتار است.
ب) توزیع پیشین جفریز را برای مدل دوجمله‌ای $f_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ محاسبه کنید.

۷. **اریبی و واریانس برآوردگر بیشترین درست‌نمایی**
الف) برای برآوردگر $\hat{\theta}$ از پارامتر θ، امید مربع خطا، MSE، به صورت $\mathbb{E}(\hat{\theta} - \theta)^2$ تعریف می‌شود. نشان دهید اگر bias و variance اریبی و واریانس تخمین‌گر $\hat{\theta}$ باشند، $\mathrm{MSE} = \mathrm{variance} + \mathrm{bias}^2$
ب) ثابت کنید برای تخمین‌گر بیشترین درست‌نمایی $\mathrm{variance} = O(1/n)\ ,\ \mathrm{bias}^2 = O(1/n^2)$

۸. **پدیده استاین در بعد یک**

الف) فرض کنید $\hat{\theta}$ تخمین‌گر دلخواهی از θ باشد. اگر $b(\theta) = \mathbb{E}[\hat{\theta}] - \theta$ اریبی $\hat{\theta}$ باشد، نشان دهید:

$$\mathrm{Var}[\hat{\theta}] \geq \frac{(1+b'(\theta))^2}{I(\theta)}$$

ب) نشان دهید در بعد یک پارادوکس استاین نداریم:
اگر $x_1, \cdots, x_n \sim \mathcal{N}(\theta, 1)$ نشان دهید با معیار امید مربع خطا، هیچ تخمین‌گری اکیداً بهتر از بیشترین درست‌نمایی $\hat{\theta} = \frac{x_1+x_2+\cdots+x_n}{n}$ نیست.

۹. **استنباط شرطی**

فرض کنید $i = 1, 2$ $x_i \sim \mathcal{N}(\mu_i, 1)$ و می‌خواهیم بردار میانگین‌ها، $\mu = (\mu_1, \mu_2)$ را برآورد کنیم. اگر بدانیم که μ در فاصله ۲ از مبدا قرار گرفته است، $\mu = (2\cos\theta, 2\sin\theta)$ و مساله برآورد θ خواهد بود. برای این کار از $\hat{\theta} = \arctan(x_2/x_1)$ استفاده می‌کنیم و قرارداد می‌کنیم که $\hat{\theta} - \theta$ از $-\pi$ تا π تغییر می‌کند.

الف) نشان دهید $\hat{\theta}$ نااریب است و امید مربع خطای آن را تا دو رقم اعشار محاسبه کنید.

ب) اگر $(x_1, x_2) = (1, 1)$ دقت $\hat{\theta}$ چقدر است؟ نظرتان درباره محاسبه خطا به شرط $r = \sqrt{x_1^2 + x_2^2}$ چیست؟!

۱۰. **تخمین ماکزیمم**

فرض کنید برای $n = 1, \cdots, N$، داریم $x_i \sim \mathcal{N}(\mu_i, 1)$ و می‌دانیم اکثر μ_iها صفر هستند. چگونه به کمک x_iها $\max_i \mu_i$ را تخمین بزنیم؟ برای پاسخ یک بار خود را آماردان فراوانی‌گرا و بار دیگر آماردان بیزگرا تصور کنید!