امتحان پایان‌ترم

آمار و احتمال مهندسی

مدرس: امیر نجفی

دانشکده مهندسی کامپیوتر – بهمن‌ماه ۱۴۰۲

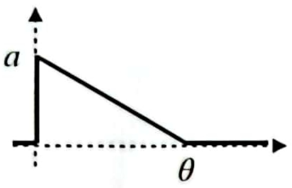
توضیحات:

* زمان کل امتحان ۱۵۰ دقیقه است. همچنین، امتحان شامل ۵ سوال است.

سوال ۱ (~~۴~~ 5 نمره):

الف) فرض کنید نمونه‌های $X_1, \ldots, X_n$ به صورت i.i.d. از یک توزیع یکنواخت در بازه $[0, \theta]$ نمونه‌گیری شده‌اند که در اینجا پارامتر θ نامعلوم فرض شده است. تخمین‌گر بیشینه درست‌نمایی برای پارامتر θ را بیابید.

ب) حال فرض کنید نمونه‌ها به جای توزیع یکنواخت از تابع چگالی احتمال زیر تولید شده باشند:



$$f_X(x) = \begin{cases} a\,(1 - x/\theta) & 0 \le x \le \theta \\ 0 & O.W. \end{cases}$$

مقدار a را بر حسب θ بیابید.

ج) هدف یافتن تخمین‌گر بیشینه درست‌نمایی برای θ است. برای این منظور، یک مسئله بهینه‌سازی طراحی کنید و پس از مشتق‌گیری، یک معادله برای نقطه بهینه آن بدست آورید (لازم نیست معادله را حل کنید، اصلاً حل بسته ندارد...).

د) **(امتیازی)** فرض کنید $M = \max_i X_i$. نشان دهید $M(1 + 1/n) \le \hat{\theta}_{\mathrm{ML}} \le 2M$.

سوال ۲ (~~۴~~ 3 نمره):

تابع مولد گشتاور (MGF) یک توزیع مانند $X \sim f_X(x)$ به صورت $M_X(t) = \mathbb{E}\left(e^{tX}\right)$ تعریف می‌شود، که t می‌تواند عددی دلخواه باشد. می‌خواهیم با استفاده از MGF یک کران بالا برای دم (tail) توزیع پیدا کنیم: یعنی مقدار $\mathbb{P}\,(X > a)$ به ازای مقادیر بزرگ برای a.

الف) اثبات کنید در صورتیکه MGF یک متغیر تصادفی نامنفی مانند X به ازای حتی یک مقدار حقیقی $t > 0$ هم وجود داشته باشد، آنگاه دم توزیع با سرعتی حداقل برابر با توزیع نمایی برحسب a کاهش خواهد یافت.

راهنمایی ۱: دقت کنید که به ازای $t > 0$، دو گزاره $X > a$ و $e^{tX} > e^{ta}$ معادل هستند.

راهنمایی ۲: نامساوی مارکوف استفاده کنید.

ب) به صورت کیفی پاسخ دهید: در این صورت، چرا دم توزیعی مانند کوشی بسیار کندتر از نمایی به سمت صفر میل می‌کند؟

سوال ۳ (۴ نمره):

یک شرکت برای شبیه‌سازی‌های مرتبط با پروژه شبکه مخابرات بی‌سیم خود، نیازمند تولید نمونه‌های i.i.d. از یک متغیر تصادفی «رایلی» در محیط پایتون است. توزیع رایلی متناظر با یک متغیر تصادفی نامنفی است و تابع چگالی به شکل زیر دارد:

$$f_X(x) = \frac{x}{\sigma^2} e^{-x^2/\left(2\sigma^2\right)}$$

که $0 < \sigma$ پارامتر آن است.
پکیج‌های موجود در رایانه‌های شرکت تنها قابلیت تولید نمونه‌های تصادفی و مستقل یکنواخت در بازه ۰ تا ۱ را دارند، که به صورت $X_1, \ldots, X_n$ نمایش داده می‌شوند. تابع g را به گونه‌ای بیابید که $g(X_1), \ldots, g(X_n)$ توزیع رایلی داشته باشند.
راهنمایی: در صورتیکه در حین حل سوال به یک انتگرال برخوردید، تغییر متغیر $u = x^2/\sigma^2$ می‌تواند موثر باشد.

سوال ۴ (~~۴~~ 3 نمره):

یک کشاورز از یک شرکت تولیدکننده سموم آفت‌کش شکایت کرده، و معتقد است سموم تولیدی آنان علاوه بر آفات، زنبورهای عسل را نیز از بین می‌برد و به وی خسارت زده است. اما شرکت مزبور این ادعا را رد می‌کند. قاضی دادگاه از یکی از دانشجویان درس آمار و احتمال این ترم دانشکده کامپیوتر برای بررسی صحت ادعای کشاورز کمک خواسته است. دانشجو در دو حالت ۱) قبل از استفاده از سم آفت‌کش، و ۲) مدتی بعد از استفاده از آن، تعداد زنبورهای عسل در هر متر مکعب از فضا را در نزدیکی کندوهای کشاورز اندازه گرفته است.

- برای حالت ۱) تعداد ۹ اندازه‌گیری مستقل به صورت $X_1, \ldots, X_9$ انجام داده که اطلاعی در مورد میانگین یا خانواده توزیع آنان ندارد. اما واریانس هر اندازه‌گیری را معلوم و برابر با ۱ فرض کرده. میانگین تجربی ۹ نمونه فوق برابر با ۱۳ زنبور بر واحد متر مکعب شده است.
- برای حالت ۲) تعداد ۱۶ اندازه‌گیری مستقل به صورت $Y_1, \ldots, Y_{16}$ انجام داده که دوباره درباره میانگین یا توزیع آنان اطلاعی ندارد. اما واریانس هر نمونه را نیز مطابق قسمت اول، برابر با ۱ فرض کرده است. این بار میانگین ۱۶ نمونه فوق برابر با ۱۰/۵ زنبور بر واحد حجم شده، که نسبت به حالت ۱ قدری کاهش را نشان می‌دهد.

ادعای کشاورز این است که **میانگین‌های** توزیع تعداد زنبورها در واحد حجم، پیش و پس از استفاده از سموم، تفاوت معنی‌دار دارند. همچنین طبق قانون، قاضی در صورتی ادعای کشاورز را تایید می‌کند که p-value مشاهدات فوق از ۱ درصد کمتر باشد.

الف) آماره لازم به منظور انجام آزمون فرض را تشکیل داده، و فرضیه باطل (Null) و جایگزین را مشخص کنید. **راهنمایی:** دقت کنید، آماره و فرضیه باطل را باید به گونه‌ای تشکیل دهید که تحت فرض باطل، حداقل میانگین و واریانس آماره مشخص باشند.
ب) یک کران بالا برای مقدار p-value مشاهدات فوق پیدا کنید. آیا ادعای کشاورز تایید می‌شود؟ **راهنمایی:** به دلیل مشخص نبودن توزیع و کم بودن تعداد مشاهدات (عدم برقراری تقریب حد مرکزی)، ناچار هستید از نامساوی «چبی‌شف» استفاده کنید...
ج) با فرض همین مشاهدات، و صرفاً اضافه کردن فرض توزیع گاوسی برای توزیع تعداد زنبورها در واحد حجم، آیا ادعای کشاورز تایید می‌شد؟ (کران دقیق برای دم توزیع گاوسی لازم نیست، از اطلاعات کلی خود در مورد شکل این توزیع برای پاسخ به قسمت ج) استفاده کنید).

سوال ۵ (~~۴~~ 5 نمره):

از یک متغیر تصادفی مانند X تعدادی نمونه i.i.d. به صورت $X_1, \ldots, X_n$ در دست هستند. در مورد خانواده توزیع X چیزی نمی‌دانیم، اما هدف ما نیز تنها تخمین مقدار احتمال $\mathbb{P}(X \geq 0)$ است و به دنبال تخمین پارامترهای بیشتری از توزیع نیستیم.
الف) با استفاده از نمونه‌ها، تخمین‌گر بیشینه درست‌نمایی را برای مقدار $\mathbb{P}(X \geq 0)$ بدست آورید.
ب) بایاس و واریانس این تخمین‌گر چقدر هستند؟ آیا این تخمین‌گر سازگار است؟ (جواب‌ها ممکن است بر حسب مقدار واقعی و نامعلوم $p = \mathbb{P}(X \geq 0)$ بیان شوند، که اشکالی ندارد).

ج) حال فرض کنید که در مورد احتمال خواسته شده اطلاعاتی موجود است که به صورت توزیع پیشین $f(p) = kp^{k-1},\ 0 \leq p \leq 1$ مدلسازی شده است ـ در اینجا k عددی بزرگ‌تر از ۱ است. در این صورت، تخمین بیشینه توزیع پسین (MAP) برای مقدار نامعلوم $p = \mathbb{P}(X \geq 0)$ چیست؟ (نشان دهید تخمین‌گر MAP در این حالت، مشابه با تخمینگر ML بوده با این تفاوت که انگار تعداد $k - 1$ عدد نامنفی به n نمونه اولیه اضافه شده‌اند).

موفق باشید.