امیرحسین باقری، محمد ابول‌نژادیان، محمدعلی حیدری

تمرین سری پنجم | آمار | مهلت ارسال: ۲۳:۵۹ ــ ۲۹ آذر ۱۴۰۰

نکات مهم

- پاسخ‌های نظری خود را را در قالب یک فایل pdf با اسم HW#_[STD-Num] در بخش مختص به خود در کوئرا آپلود کنید. پاسخ‌های عملی را نیز با نامی مشابه، ولی با فرمت zip در بخش مختص خود آپلود کنید.
- تنها سوالات ستاره‌دار تمرین، نیاز به تحویل دارند. نوشتن پاسخ‌های نظری، هم به صورت تایپی و هم دست‌نویس، مقبول است. پاسخ‌های عملی باید در زبان R نوشته شوند.
- سوالات پرسیده شده در خصوص تمرین در شبکه‌های اجتماعی، به هیچ عنوان پاسخ داده نخواهند شد؛ تنها مکان مجاز رفع اشکال در خصوص تمرین، بخش پرسش‌ها و پاسخ‌ها در کوئرا است.
- زمان تحویل تمرین، به هیچ عنوان تمدید نخواهد شد، بنابراین لازم است که زمان خود را برای انجام تمرین مدیریت کرده و آن را به روزهای پایانی موکول نکنید.
- امکان بارگذاری تمرین در کوئرا تا ۷۲ ساعت پس از ددلاین تمرین وجود دارد، اما به ازای هر ساعت تاخیر، یک درصد از نمره‌ی نهایی تمرین را از دست خواهید داد. دقیقه‌ها و ثانیه‌ها، رو به بالا گرد خواهند شد؛ مثلا، یک ساعت و نیم تاخیر، معادل دو ساعت تاخیر محسوب می‌شود.
- در طول ترم، ۲۴۰ ساعت کوپن تاخیر خواهید داشت و با استفاده از آن‌ها، می‌توانید بدون کسر نمره، از تاخیرها استفاده کنید. جریمه‌ی تاخیرها، از تمرینی محاسبه می‌شوند که در آن، کوپن‌ها به اتمام رسیده باشند. نمره‌ی امتیازی برای اشخاصی که مجموع تاخیرهای ایشان در کل ترم، کم‌تر از ۲۴۰ ساعت باشد، به هیچ عنوان در نظر گرفته نمی‌شود.
- مشورت در تمرین‌ها مجاز است و توصیه هم می‌شود، اما هر دانش‌جو موظف است تمرین را به تنهایی انجام دهد و راه‌حل نهایی ارسال شده، باید توسط خود دانش‌جو نوشته شده باشد. در صورت کشف اولین مورد تقلب هر دانش‌جو، نمره‌ی همان تمرین وی، صفر در نظر گرفته شده و در صورت کشف دومین مورد تقلب هر دانش‌جو، منفی نمره‌ی کل تمرین‌ها به وی تعلق خواهد گرفت. برای کسب اطلاعات بیش‌تر در خصوص آیین‌نامه‌ی مشورت و تقلب، می‌توانید به بخش مربوطه در ویکی دانشکده مراجعه کنید. لازم به ذکر است که این جرایم به هیچ عنوان بخشیده نخواهند شد.

سوالات نظری

مسئله‌ی ۱. احتمال و آمار؟

الف) در دو خط، تفاوت کلی علم احتمال و آمار را توضیح دهید.

ب) یکی از تفاوت‌های کلی دیدگاه‌های frequentist و Bayesian را به اختصار توضیح دهید.

مسئله‌ی ۲. یک‌نواخت *

فرض کنید توزیع $uniform(a,b)$ داده شده است.

الف) ۳ داده‌ی ۱۰ و ۸ و ۱۲/۵ از این توزیع انتخاب شده‌اند. maximum likelihood را برای پارامترهای a و b تخمین بزنید.

ب) اگر n داده از این توزیع به صورت $x_۱, x_۲, ..., x_n$ انتخاب شده باشند، maximum likelihood را برای پارامترهای a و b به دست آورید.

مسئله‌ی ۳. چرنوبیل *

منبعی رادیواکتیو در چرنوبیل وجود دارد که در هر زمانی که اندازه‌گیری می‌شود، K فوتون از خود ساطع می‌کند. ما فرض می‌کنیم که که K دارای توزیع زیر است:

$$p_K(k;\theta) = c(\theta)e^{-k\theta} \qquad k = ۰, ۱, ۲, ...$$

به طوری که θ معکوس دمای منبع بوده و $c(\theta)$ کمیتی به نام «ضریب نرمالیزاسیون» است . هم‌چنین، فرض می‌کنیم که فوتون‌های ساطع شده از منبع، هر بار مستقل از یک‌دیگر هستند. ما می‌خواهیم دمای منبع را با اندازه‌گیری‌های پی در پی فوتون‌های ساطع شده، تخمین بزنیم.

الف) ضریب نرمالیزاسیون $c(\theta)$ را بیابید.

ب) تخمین ML برای دمای منبع $T = \frac{۱}{\theta}$ را بر اساس $K_۱, K_۲, ...$ (که K_i تعداد فوتون های ساطع شده در اندازه‌گیری i ام است) را بیابید.

مسئله‌ی ۴. بازه‌های شکست *

یک متغیر برنولی با احتمال موفقیت p را در نظر بگیرید که از مقدار p مطلع نیستیم. حال متغیر T را به صورت زیر برای هر موفقیت تعریف می‌کنیم:
T_i برابر تعداد شکست های بین موفقیت $T_{i-۱}$ و T_i، با احتساب موفقیت K ام است؛ به عبارت دیگر،

$$T_1 = Y_1 \ , \ T_k = Y_k - Y_{k-1}, \ k = 1, 2, 3, ...$$

که متغیر Y_K برابر شماره‌ی آزمایشی است که K امین موفقیت را در پی داشته است. تعداد کل آزمایش‌ها به ما داده نشده است، بلکه تنها مجموعه‌ی T را در اختیار داریم که برابر $T = \{T_۱, T_۲, ..., T_k\}$ است.

الف) یک تخمین‌گر بیشینه‌ی درست‌نمایی برای پارامتر P به دست آورید. (دقت کنید که راه‌حل شما باید به صورت maximum likelihood باشد)

ب) با فرض دانستن p به نظر شما توزیع T_i به چه توزیع‌ای شبیه است ($PMF_{T|p}$)؟ میانگین و واریانس این توزیع را به دست آورید. سپس با استفاده از توزیع T و با فرض ندانستن p و هم‌چنین بدون هیچ پیش‌فرضی از داده‌های به دست آمده (یعنی فرض کنید که هیچ داده‌ای نیز به شما داده نشده است)، تابع PMF_T را به دست آورید.

ج) نشان دهید که رابطه‌ی زیر برقرار است (p^* برابر مقدار واقعی پارامتر توزیع است):

$$\forall\epsilon > ۰ \;:\; {}_{k\rightarrow\infty}P(|\tfrac{۱}{\hat{P}} - \tfrac{۱}{p^*}| > \epsilon) = ۰$$

د) فرض کنید که $p^* < ۰/۵$ است. به کمک نامساوی چبی‌شف، حال یک کران پایین برای K بیابید به صورتی که رابطه‌ی زیر برقرار باشد:

$$P(|\tfrac{۱}{\hat{P}} - \tfrac{۱}{p^*}| \leqslant ۰/۱) \geqslant ۰/۹۵$$

مسئله‌ی ۵. نقطه‌ای

تا این‌جا با تخمین‌هایی آشنا شده‌اید که یک پارامتر تابع را تخمین می‌زنند؛ حال می‌خواهیم با تخمین دیگری آشنا شویم که به صورت نقطه‌ای توزیع را تخمین می‌زند.

فرض کنید یک تابع Q و هم‌چنین یک متغیر h_q نیز به شما داده شده است که دارای ویژگی‌های زیر هستند:

$$\begin{cases} \forall u \;:\; ۰ \leqslant Q(u) \\ \int_R Q(x)dx = ۱ \\ {}_{N\longrightarrow\infty}Q(x_i) = \delta(x_i) \\ {}_{N\longrightarrow\infty}\frac{۱}{h_q}Q(\frac{a-x_i}{h_q}) = \delta(a-x_i) \end{cases}$$

$$\begin{cases} u < \frac{۱}{۲} \;:\; Q(u) = ۱ \\ o.w \;:\; Q(u) = ۰ \end{cases}$$

حال فرض کنید که مقدار تابع را به صورت زیر تخمین می‌زنیم (و تعداد N داده به ما داده شده است):

$$f(\hat{a}) = \frac{۱}{N*h_q}\sum_{q=۱}^{N} Q(\frac{a-x_q}{h_q})$$

در حقیقت نسبت تعداد داده‌هایی که در فاصله‌ای کمتر از $\frac{h_q}{۲}$ هستند، به تعداد کل داده‌ها را برای این تخمین‌گر استفاده می‌کنیم (البته تقسیم بر طول بازه نیز انجام می‌شود که یک تابع چگالی احتمال به دست آید).

فرض کنید تابع $f^*(a)$ به شما داده شده است. ثابت کنید که این تخمین‌گر، بایاس ندارد. (دقت کنید که کانولوشن یک تابع با تابع دلتای دیراک، برابر خود تابع است)

مسئله‌ی ۶. آماردان *

شما به عنوان یک آماردان در شرکتی مشغول به کار شده‌اید. طراح یک بازی، به شما مراجعه کرده و می‌خواهد بداند که برد افراد در بازی شرطی‌اش به چه صورت است. شما در بررسی میزان برد افراد در این بازی، به متغیرهای

$X_1, X_2, X_3, \ldots$ که متغیرهایی iid با توزیع یکنواخت $U(-\theta, \theta)$ هستند، رسیدید ($\theta > 0$).

الف) تخمین بیشینه‌ی درست‌نمایی $\hat{\Theta}$ را بیابید.

ب) توزیع دقیق $\hat{\Theta}$ را به دست آورید.

ج) آیا تخمین‌گر به دست آمده بدون bias است؟ در صورت وجود bias یک تخمین‌گر بدون bias ارائه دهید.

مسئله‌ی ۷. تاس‌های نامتقارن *

یک تاس k وجهی به شما داده شده است؛ هم‌چنین یک مجموعه‌ی نمونه از نتایج پرتاب این تاس در اختیار داریم. تعداد باری که وجه i ام تاس ظاهر شده، m_i است و می‌دانیم $\sum_{i=1}^{k} m_i = N$. احتمال آمدن وجه i ام تاس نیز با μ_i نمایش داده می‌شود و هم‌چنین $X(x) = i$ نشان دهنده مقدار متغیر تصادفی وجه رو آمده از تاس است (به عنوان مثال $X(\omega) = 1$ یعنی وجه اول تاس آمده است). تعداد کل پرتاب‌های انجام شده، N تا است.

الف) فرض کنید که تاس، دو وجهی است؛ یک تخمین‌گر بیشینه‌ی درست نمایی برای احتمال رو آمدن هر وجه پیدا کنید. (دقت کنید که راه حل شما حتما باید شامل maximum likelihood باشد و به این‌که از نظر فیزیکی آیا تاس ۲ وجهی وجود خارجی دارد یا خیر نیز توجه نداشته باشید)

ب) این بار، فرض کنید که k وجه داریم و مسئله را به صورت کلی حل کنید، یعنی یک تخمین‌گر بیشینه‌ی درست نمایی برای هر یک از μ_k به دست آورید. (دقت کنید که راه‌حل شما باید شامل maximum likelihood باشد و راه‌حل‌هایی نظیر استقرا مجاز نیستند)

سوالات عملی *

سوالات عملی که تحویل آن‌ها اجباری است، به صورت یک ژوپیتر نوت‌بوک در کوئرای درس قرار داده شده‌اند. لازم است این ژوپیتر نوت‌بوک را طبق دستورالعمل‌های نوشته شده در آن، تکمیل کرده و در کوئرا آن را آپلود کنید.

موفق باشید! :)