به نام خدا

دانشگاه صنعتی شریف

دانشکده مهندسی برق

گروه دکتر یاسایی ـ آمار و احتمال مهندسی

نیم سال اول ۱۴۰۱ ـ ۱۴۰۰

تمرین عملی سری دوم

لطفاً به نکات زیر توجه بفرمایید:

۱. نتایج و پاسخ های خود را در یک فایل با فرمت zip به نام HW2-Name-StudentNumber در سایت cw قرار دهید.

۲. کسب نمره کامل در هر سؤال مستلزم تحویل **کدها** و **توضیحات** می‌باشد.

۳. برای سؤالات، باید روشی که استفاده کرده‌اید را توضیح و نتایجی که گرفته‌اید را ارائه دهید. این توضیحات می‌تواند در یک فایل pdf و یا در یک فایل ipynb باشد.

۴. کدهای خود را خوانا بنویسید و کامنت‌گذاری کنید. در plot های خود عنوان، label و خطکشی‌های مناسب را اضافه کنید.

۵. ابهام یا اشکالات خود را مي توانيد از طريق Amirhosein_Javadi@ يا Javadiamirhosein.2000@gmail.com مطرح نمایید.

۶. کدهای شما تماماً باید توسط خودتان نوشته شده باشند. هرگونه استفاده از کد دیگران به هر شکل ممکن، تقلب محسوب می‌شود و نمره تمرین کامپیوتری جاری صفر خواهد شد. پس در هیچ صورت کدهای خود را برای دیگران ارسال نکنید.

۷. مهلت تحویل:

۱ حرکت براونی و توزیع نرمال

در این سوال قصد داریم با توزیع نرمال و حرکت براونی بیشتر آشنا شویم. فرض کنید ذره‌ای در هر واحد زمانی δ ثانیه‌ای به اندازه‌ی ϵ با احتمال برابر به سمت راست یا چپ حرکت می‌کند و هدف ما پیدا کردن مکان ذره در زمان t، B_t، است.

۱. زمان را با نرخ 10000 سمپل بر ثانیه نمونه برداری کنید. فاصله‌ی بین هر دو نمونه در واقع همان δ مسئله است و در کدام از این بازه‌های زمانی به اندازه‌ی $\epsilon = \frac{2}{100}$ حرکت کنید.

۲. حال میتوانید تابعی تعریف کنید که خروجی آن متغیر تصادفی از مکان ذره در زمان $t = 1$ باشد. این تابع را 10000 بار اجرا کنید و در یک نمودار هیستوگرامی از B_1 رسم کنید. انتظار دارید این هیستوگرام متناظر با چگالی احتمال چه متغیر تصادفی با چه پارامترهایی باشد؟ تابع چگالی احتمال این متغیر تصادفی را هم در همین نمودار رسم کنید.

۳. قسمت دوم را برای B_2، B_3، B_4 تکرار کنید.

۲ تخمین پارامتر p متغیر تصادفی برنولی

در این سوال قصد داریم با کمک تخمینگر Maximum Likelihood پارامتر p یک متغیر تصادفی برنولی را حدس بزنیم.

۱. تابعی بنویسید که در هر اجرای آن، عدد تصادفی p بین 0 تا 1 به صورت تصادفی یکنواخت انتخاب کند. سپس $n = 1000$ نمونه از متغیر تصادفی برنولی با پارامتر p بردارید و در آرایه‌ای با نام Samples ذخیره کنید. آرایه‌ای به نام ML_Estimator به طول n تعریف کنید و درایه‌ی iام آن را برابر خروجی تخمینگر Maximum Likelihood در حالتی که فقط i درایه‌ی اول آرایه‌ی Samples را میدانید، قرار دهید. آرایه‌ی خطای تخیمنگر را برابر قدر مطلق اختلاف هر درایه‌ی ML_Estimator و p قرار دهید و این آرایه را به عنوان خروجی تابع تعیین کنید. در یک نمودار خطای تخمینگر را رسم کنید.

۲. تابع بالا را 1000 بار تکرار کنید و خطای میانگین در این 1000 تکرار را به دست بیاورید. در یک نمودار خطای میانگین تخمینگر را رسم کنید.

۳ تخمین پارامترهای μ و σ متغیر تصادفی نرمال

در این سوال قصد داریم با کمک تخمینگر Maximum Likelihood پارامترها μ و σ یک متغیر تصادفی نرمال را حدس بزنیم. تابعی بنویسید که در هر اجرای آن، عدد تصادفی μ و σ را به صورت تصادفی یکنواخت از بازه‌ی $[-2, 2]$ و $[0, 5]$ انتخاب کند و این دو عدد را چاپ کند. سپس $n = 1000$ نمونه از متغیر تصادفی نمایی $\mathbb{N}(\mu, \sigma^2)$ بردارید و در آرایه‌ای با نام Samples ذخیره کنید. آرایه‌ای به نام Mu_Estimator و Sigma_Estimator به طول n تعریف کنید و درایه‌ی iام آن را برابر خروجی تخمینگر Maximum Likelihood در حالتی که فقط i درایه‌ی اول آرایه‌ی Samples را میدانید، قرار دهید. دقت کنید که از μ و σ نمیتوانید برای این تخمین استفاده کنید. آرایه‌ی خطای تخیمنگر μ را برابر قدر مطلق اختلاف هر درایه‌ی Mu_Estimator و μ و آرایه‌ی خطای تخیمنگر σ را برابر قدر مطلق اختلاف هر درایه‌ی Sigma_Estimator و σ قرار دهید. بهترین تخمین از μ و σ را چاپ کنید. این دو آرایه را به عنوان خروجی‌های تابع تعیین کنید. در دو نمودار این دو خطای تخمینگر را رسم کنید.

۴ تست فرضیه

در این سوال قصد داریم با مفهوم hypothesis testing بیشتر آشنا شویم. فرض کنید دو متغیر تصادفی نرمال H_0 و H_1 به صورت زیر داریم.

$$H_0 \sim \mathbb{N}(\mu = -1, \sigma = 5) \tag{۱}$$

$$H_1 \sim \mathbb{N}(\mu = 1, \sigma = 5) \tag{۲}$$

۱. تابع چگالی احتمال H_0 و H_1 را در یک نمودار رسم کنید.

۲. همان طور که در درس دیدیم برای حدس زدن مدل بین H_0 و H_1 با داشتن نمونه‌ی x میتوانیم از تست زیر استفاده کنیم:

$$f_{H_1} \underset{H_0}{\overset{H_1}{\gtrless}} f_{H_0} \Leftrightarrow (x+1)^2 \underset{H_0}{\overset{H_1}{\gtrless}} (x-1)^2 \Leftrightarrow 2x \underset{H_0}{\overset{H_1}{\gtrless}} -2x \tag{۳}$$

در این سوال قصد داریم بررسی کنیم آیا با نمونه‌های بیشتر میتوانیم مدل‌های H_0 و H_1 را بهتر از هم جدا کنیم یا نه. به همین منظور با داشته داده‌های $x_{i=1:n}$ تست ۳ را به تست زیر تعمیم می‌دهیم.

$$f_{H_1} \underset{H_0}{\overset{H_1}{\gtrless}} f_{H_0} \Leftrightarrow \sum_{i=1}^{n}(x_i+1)^2 \underset{H_0}{\overset{H_1}{\gtrless}} \sum_{i=1}^{n}(x_i-1)^2 \Leftrightarrow \sum_{i=1}^{n} 2x_i \underset{H_0}{\overset{H_1}{\gtrless}} \sum_{i=1}^{n} -2x_i \tag{۴}$$

برای این کار تابعی بنویسید که یک آرایه به نام Indicator شامل $n = 100$ نمونه از متغیر تصادفی برنولی با پارامتر $\frac{1}{2}$ تولید کند. حال n آزمایش انجام می‌دهیم و در هر آزمایش سعی داریم درایه‌ی iام آرایه‌ی Indicator را حدس بزنیم. همچنین داریم:

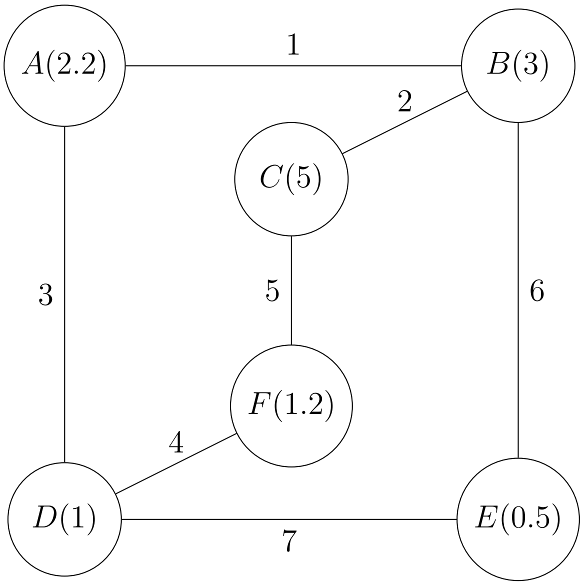
$$H^i = \begin{cases} H_1 & Indicator[i] = 1 \\ H_0 & Indicator[i] = 0 \end{cases} \tag{۵}$$

پس در هر آزمایش H^i را تعیین کنید و برای هر H^i تست زیر را 10000 بار تکرار کنید.
از مدل مربوطه i نمونه بگیرید. حال تست ۴ را اجرا کنید و تخمین خود را از درایه‌ی iام آرایه‌ی Indicator به دست بیاورید و در یک آرایه به نام Estimation ذخیره کنید. خطای تخمین برابر قدرمطلق اختلاف بین Indicator و Estimation است.

در یک نمودار میانگین خطا در 10000 تست را به ازای i نمونه رسم کنید.

۵ جمع رئوس

گراف زیر را در نظر بگیرید. هر کدام از رئوس این گراف یک عدد مخصوص دارد و هدف ما پیدا کردن مجموع عددهای رئوس این گراف در هر راس از گراف است. در اینجا هر راس می‌تواند با راس همسایه تبادل اطلاعات کند ولی رئوس حافظه محدودی دارند و می‌توانند فقط تعداد محدودی عدد در خود ذخیره کنند. یال‌ها به دلخواه شماره‌گذاری شده‌اند و عدد هر راس داخل آن در پرانتز نوشته شده‌است.



یک ایده برای تخمین مجموع مقادیر رئوس این است که هر راس نمونه‌ی تصادفی از متغیر تصادفی نمایی با با نرخ عدد آن راس تولید کند و نمونه‌ی خود را با راس همسایه تبادل کند و پس از آن دو راس همسایه نمونه‌ی خود را با مینیمم نمونه‌ی خود و همسایه جایگزین کنند. پس از چند مرحله همه رئوس مینیمم نمونه‌های تصادفی گره‌های مختلف را میبینند و الگوریتم همگرا می‌شوند و گره‌ها میتوانند مقدار مینیمم را ذخیره کنند. حال می‌دانیم مینیمم n متغیر تصادفی نمایی با نرخ‌های λ_i، یک متغیر تصادفی نمایی با نرخ $\sum_{i=1}^{n} \lambda_i$ است. پس اگر تعداد مناسبی این الگوریتم را تکرار کنیم، همه رئوس به چند نمونه iid از متغیر نمایی با پارامتر مجموع مطلوب ما دسترسی دارند و می‌توانند با تخمین Maximum Likelihood پارامتر نرخ متغیر نمایی را تخمین بزنند.

۱. می‌توانید یال‌ها را شماره‌گذاری کنید یا از شماره‌گذاری پیشنهادی استفاده کنید. در هر راند به ترتیب، یالی را انتخاب کنید و بین دو نمونه‌ی رئوس دو سر یال مینیمم‌گیری کنید و در صورت نیاز نمونه‌ی رئوس را به روز رسانی کنید.

۲. حداقل تعداد راند مورد نیاز برای این که مطمئن شویم نمونه‌ی مینیمم به همه‌ی گره‌ها رسیده است چه قدر است؟

۳. الگوریتم بالا را اجرا کنید و مجموع اعداد رئوس را با حداکثر خطای 0.1 به دست بیاورید.