810102480

گزارش پروژه 2 آمار و احتمال

بیتا نوروزی

سوال 1: ① ابتدا کتابخانه pandas را import می‌کنیم و با تابع read_csv فایل را خوانده ستون‌های metro و brt را جداگانه list می‌کنیم

② چون تعداد آزمایش‌ها بسیار زیاد و هر پیشامد امکان رخ دادنش در بازه زمانی خاص کوچک کم است، و همچنین از روی نمودار در متغیر X و Y مشهود است که هر دو دارای توزیع پواسون می‌باشند.

حال برای محاسبه پارامتر λ، به تقریب به علت زیاد بودن آزمایش‌ها، λ هر متغیر را برابر میانگین ستون متناظر قرار دادیم.

$X \sim Poi(\lambda_x)$ $Y \sim Poi(\lambda_y)$ $\lambda_x = \mu_x$ $\lambda_y = \mu_y$ $\xrightarrow{\text{با محاسبه در Python}}$ $\lambda_x = 3.5316$

$\lambda_y = 2.0636$

③ تابع چگالی metro با نمودار hist رسم می‌کنیم.

④ برای توزیع X و Y مستقیماً از تابع Poisson که در کتابخانه scipy موجود است استفاده کرده و pmf آن را با plt رسم می‌کنیم. طبق این نمودار و نمودار قبلی تقریباً یکسان اند و نقطه پیک و دیگر نقطه‌ها نیز تقریباً درست عمل شوند

⑤ X و Y هر دو مستقل از هم و توزیع یکسانی دارند. بنابراین Z که حاصل جمع X و Y می‌باشد نیز دارای توزیع پواسون و λ نیز از حاصل جمع λ_x و λ_y به دست می‌آید. $Z = X+Y \to Z \sim Poi(\lambda_x+\lambda_y) \to Z \sim Poi(5.5952)$. آنگاه metro, brt را نیز مستقیم جمع کنیم، آن را با poisson.pmf با پارامتر جدید دهیم، نمودار هر دو تقریباً روند یکسانی خواهند داشت و درست است.

⑥ توزیع W را ساده کرده و با Python می‌دهیم.

$$W \sim P(X|X+Y=n) = \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)} = \frac{P(X=k)P(Y=n-k)}{P(Z=n)} = \frac{\cancel{e^{-\lambda_x}}\frac{(\lambda_x)^k}{k!}\cdot\cancel{e^{-\lambda_y}}\frac{(\lambda_y)^{n-k}}{(n-k)!}}{\cancel{e^{-(\lambda_x+\lambda_y)}}\frac{(\lambda_x+\lambda_y)^n}{n!}}$$

$$= \frac{n!}{k!(n-k)!}\frac{(\lambda_x)^k(\lambda_y)^{n-k}}{(\lambda_x+\lambda_y)^k(\lambda_x+\lambda_y)^{n-k}} = \binom{n}{k}\left(\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^k\left(\frac{\lambda_y}{\lambda_x+\lambda_y}\right)^{n-k}$$

$$\Longrightarrow W \sim Bin\left(n, \frac{\lambda_x}{\lambda_x+\lambda_y}\right)$$

⑦ برای رسم pmf W با مقدار n=8 ← $Bin(8, \frac{3.5316}{5.5952})$. این را رسم می‌کنیم با تابع binom.pmf در scipy

⑧ لیست جدیدی می‌سازیم که مجموع metro و brt را در هر نقطه زمانی داشته باشیم. حال طبق خواسته سوال بازه‌هایی که در لیست جدید مجموع metro و brt 8 می‌باشد را پیدا کرده و در لیست metro در آن بازه زمانی متناظر تعداد حرکات metro را به دست می‌آوریم و ذخیره می‌کنیم. حال برای این مقادیر ذخیره شده، تابع چگالی را رسم کرده و مشاهده می‌کنیم که دوباره روند یکسانی دارند.

سوال 2: ① مقادیر n و k را به عنوان ورودی می‌دهیم و آزمایش را k بار انجام می‌دهیم. در هر آزمایش، عددی تصادفی از 1 تا n را انتخاب می‌کنیم و بررسی می‌کنیم که اگر عدد قبلاً مشاهده نشده باشد آن را به لیست اضافه می‌کنیم و تعداد آزمایش‌ها را هر بار یکی افزایش می‌دهیم. در آخر مجموع تمام تعداد آزمایش‌ها را در k بار آزمایش حساب کرده و برای میانگین مجموع آنها را بر k تقسیم می‌کنیم.

② چون اعداد تولید شده random می‌باشند با انجام آزمایش‌ها و مشاهده میانگین با سه مقدار k داده شده یعنی 10، 100، 1000 می‌بینیم که هر چقدر k افزایش یابد عدد ما به 29 نزدیکتر می‌شود و می‌توان گفت مقدار میانگین محاسبه شده با افزایش k، به 29.3 همگرا می‌باشد.

ادامه ② و ③ متغیرهای X_i همگی آزمایش‌هایی مستقل از هم و تا اولین موفقیت ادامه پیدا می‌کنند ← توزیع هندسی

همچنین چون تعداد اعداد مشاهده شده در حال افزایش است، با افزایش i، احتمال مشاهده عدد i‌ام برای اولین بار کم می‌شود برای iهای متفاوت متفاوت است.

$X_i \sim Geo(p_i) \qquad p_i = 1-\frac{i-1}{n} \rightarrow X_i \sim Geo\left(1-\frac{i-1}{n}\right) \rightarrow P_{X_i}(k) = p_i(1-p_i)^{k-1}$

محاسبه MGF هر X_i:

$$\Phi_{X_i}(s) = E(e^{sX_i}) = \sum_{k=1}^{\infty} e^{sk} P_{X_i}(k) = \sum_{k=1}^{\infty} e^{sk} p_i (1-p_i)^{k-1} = \left(1-\frac{i-1}{n}\right) \sum_{k=1}^{\infty} \left(\frac{i-1}{n}\right)^{k-1} e^{sk}$$

$$= \left(1-\frac{i-1}{n}\right)\left(\frac{e^s}{1-\left(\frac{i-1}{n}\right)e^s}\right) \Rightarrow \Phi_{X_i}(s) = \frac{p_i e^s}{1-(1-p_i)e^s} \qquad p_i = 1-\frac{i-1}{n}$$

④ $X = \sum_{i=1}^{n} X_i$ و همچنین MGF برای مجموع متغیر تصادفی مستقل ← حاصل‌ضرب MGF های هر کدام از X_i

در نتیجه $\Phi_X(s) = \Phi_{X_1}(s)\,\Phi_{X_2}(s) \cdots\cdots \Phi_{X_n}(s)$ پس تمامی Φ_{X_i} های بالا را در یک لیست می‌بریم و با ضرب کردن آن‌ها به $\Phi_X(s)$ می‌رسیم.

⑤ می‌دانیم که برای $E(X)$ می‌توانیم از $\Phi_X(s)$ مشتق بگیریم و در نقطه ۰ آن را محاسبه کنیم. پس $\Phi_X(s)$ بالا را مشتق و سپس s را ۰ قرار داده، با مشاهده $E(X)$ به دست آمده: 29,2896 می‌بینیم که دقیقاً مانند بخش 2 عددی بسیار نزدیک به 29,3 در آن حاصل می‌شود.

سوال 3: ① Lable 201 و 202 یعنی ردیف‌های 200, 201 را با تابع drop حذف می‌کنیم.

② عملاً تابع پله‌ای $u(t-128)$ را اعمال می‌کنیم. به جز ستون Lable تمامی مقادیر Pixel ها را چک کرده و آن از 128 کوچکتر بود ۰ (خاموش) و آن بزرگتر مساوی 128 بود آن را ۱ (روشن) می‌کنیم.

③ با استفاده از تابع random عددی تصادفی بین ۰، ۱۹۹ برای انتخاب ردیف انتخاب می‌کنیم و Pixel های آن که 1 تا 784 است در یک ماتریس 28×28 قرار می‌دهیم. با نمایش این آرایه دوبعدی با تابع ذکر شده، با هر بار اجرا، یکی از اعداد Lable ها یعنی 8 یا 9 به طور گرافیکی نمایش داده می‌شود که البته برای هر ردیف مدل و font منحصر به فردی دارد.