گزارش کار پروژه دوم آمار و احتمال

البرز محمودیان
۸۱۰۱۰۱۰۱۴

سوال ۱) ۱- فایل csv را می‌خوانیم و ستون‌های مربوط به metro و brtها را ذخیره می‌کنیم

۲- X و Y هر دو دارای توزیع پواسون هستند (از روی نمودارهایشان) و همچنین به دلیل اینکه بازه‌های زیاد و با طول کم داریم که هر پیشامد در این بازه‌های کوچک حداکثر یکبار رخ می‌دهد. و λ های هر توزیع را با میانگین هر ستون به دست می‌آوریم ← $X \sim poi(\lambda_x = \mu_x)$ و $Y \sim poi(\lambda_y = \mu_y)$

۳- تابع چگالی مربوط به ستون مترو را به صورت هیستوگرام رسم می‌کنیم

۴- با کتابخانه spicy ← pmf آنها را با plot رسم می‌کنیم و با نمودار بخش قبل مقایسه می‌کنیم که تقریبا یکسان هستند و پراکندگی تقریبا مشابهی دارند

۵- از آنجایی که X و Y مستقل و هر دو دارای توزیع پواسون هستند پس

$$Z = X + Y \rightarrow Z \sim poi(\lambda_x + \lambda_y)$$

و اگر pmf ، Z را رسم کنیم (با پارامتر $\lambda_x + \lambda_y$) ، همینطور هیستوگرام مجموع دو ستون metro و brtها را حساب کنیم مشاهده می‌کنیم که به توزیع‌های یکسانی خواهیم رسید

۴-

$$W \sim P(X \mid X+Y=n) = \frac{P(X=k, Y=n-k)}{P(X+Y=n)}$$

$$= \frac{P(X=k)\,P(Y=n-k)}{P(X+Y=n)} = \frac{e^{-\lambda_x}\frac{(\lambda_x)^k}{k!}\cdot e^{-\lambda_y}\frac{(\lambda_y)^{n-k}}{(n-k)!}}{e^{-(\lambda_x+\lambda_y)}\frac{(\lambda_x+\lambda_y)^n}{n!}}$$

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

$$= \binom{n}{k}\left(\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^k\left(\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^{n-k} \Rightarrow W \sim Bin\left(n, \frac{\lambda_x}{\lambda_x+\lambda_y}\right)$$

۷) با تابع binom.pmf ← pmf ، W را رسم می کنیم

۸) ابتدا لیستی می سازیم تا در هر بازه زمانی مجموع متردو و brt را نشان دهد
و زمان هایی را که این مجموع برابر n شده را پیدا می کنیم و در یک لیست جدید
تعداد متردو های این زمان ها را نگه می داریم و تابع جرمی آن را رسم می کنیم
و مشاهده می کنیم که دوباره دو نمودار روند یکسانی دارند

سوال ۲) ۱) تابع مونته کارلو را پیاده سازی می‌کنیم و هر بار که کارت جدیدی رو می‌کنیم به count یکی اضافه می‌کنیم و در آخر تعداد کارت هایی که لازم بود رو شوند تا n نوع کالا ابرک مشاهده شوند را تقسیم بر k می‌کنیم و return می‌کنیم

۲) مشاهده می‌کنیم که هر چقدر k را افزایش دهیم میانگین ما به عدد ۲۹٫۳ همگرا می‌شود.

۳) X_i ها همگی مستقل و توزیع هندسی دارند

$$X_i \sim Geo(P_i) \quad P_i = 1 - \frac{i-1}{n} \quad P_{X_i}(k) = P_i(1-P_i)^{k-1}$$

$$\phi_{X_i}(s) = \sum_{k=1}^{\infty} e^{sk} P_{X_i}(k) = \left(1 - \frac{i-1}{n}\right) \sum_{k=1}^{\infty} e^{ks} \left(\frac{i-1}{n}\right)^{k-1}$$

$$= \left(1 - \frac{i-1}{n}\right)\left(\frac{e^s}{1-\left(\frac{i-1}{n}\right)e^s}\right) = \frac{P_i e^s}{1-(1-P_i)e^s}$$

۴) می‌دانیم MGF مجموع چند متغیر تصادفی مستقل برابر ضرب MGF هر کدام از آن X_i هاست

$$\Rightarrow \phi_X(s) = \prod_{i=1}^{n} \phi_{X_i}(s)$$

۵) از $\phi_X(s)$ که در قسمت قبل به دست آوردیم یکبار مشتق می‌گیریم و $s=0$ تا به $E[X]$ برسیم که به عدد ۲۹٫۲۸ می‌رسیم که مانند قسمت ۲ نتیجه به دست آمده بسیار نزدیک به ۲۹٫۳ است

subject : date :

سوال ۳) ۱) ردیف های ۲۰۰ و ۲۰۱ ام را ذخیره می کنیم و با drop آنها را حذف می کنیم

۲) تمامی مقادیر را چک می کنیم که اگر کوچکتر از ۱۲۸ بود آن را خاموش (0) وگرنه روشن (1) می کنیم

۳) با تابع random.randint یک ردیف را به طور رندوم انتخاب می کنیم و pixel های آن را در یک ماتریس ۲۸ در ۲۸ نگه می داریم که با نمایش این ماتریس عدد ۸ یا ۹ به صورت گرافیکی نمایش داده می شود که البته label هر ردیف font متفاوتی دارد.