آمار و احتمال

گزارش کار پروژه 1

کسری کاشانی نژاد

810101490

---

1) برای مقادیر میانگین، در هر دو حالت تئوری و عملی، نمودارها تا حد خوبی به هم نزدیک اند و انگار روی هم افتاده اند. اما با کمی دقت و zoom می توان فهمید که کمی با هم فاصله دارند. به طوری که در ابتدا میانگین عملی کمی بالاتر از میانگین تئوری قرار گرفته و سپس از یک جایی به بعد، میانگین تئوری بالاتر قرار می گیرد.

برای مقادیر واریانس، در همان نگاه اول تفاوت دو نمودار در هر حالت مشهود است. به طوری که نمودار حالت تئوری مانند سهمی رو به پایین مثل

می باشد و خیلی از دور بالا و پایین ندارد. اما نمودار حالت عملی نامنظم است و بالا و پایین دارد و مثل

می باشد.

در هر دو مقدار، حالت تئوری با فرمول، چون می توان اثبات کرد، دقیق تر و منظم تر است.

در تابع binomial، m نمونه و n آزمایش با احتمال موفقیت p می گیریم و به تابع random.choice می دهیم و به ما آرایه ای با n×m عضو شامل 0 و 1 می دهد. آرایه را به ماتریس تبدیل می کنیم سپس در هر نمونه 1 ها را جمع می کنیم تا تعداد موفقیت های آن نمونه را بیابیم و در آخر این تعداد ها را برمی گردانیم. در حالات تئوری، با فرمول های np و $np(1-p)$ که p از 0.01 تا 1 هست و در حالات عملی نیز این مقادیر p را هر بار به تابع binomial می دهیم.

---

2) در اینجا $n=250$, $p=0.008$ می باشد. با هر سه توزیع دو جمله ای و پوآسون و نرمال، احتمال اینکه روزانه x نفر در تصادفات بمیرند را حساب می کنیم به طوری که x می تواند از 0 تا n یعنی 0 تا 250 باشد. پارامترهای مربوط به هر توزیع را به آن می دهیم. برای توزیع های دو جمله ای، پوآسون و نرمال به ترتیب به pmf آن n و p، به pmf آن $\lambda=np$، به pdf آن (pdf چون پیوسته است.) $\mu=np$ و $\sigma^2=np(1-p)$ را می دهیم و در آخر هر سه توزیع را رسم می کنیم.

مشاهده می شود که نمودار توزیع پوآسون با خطای کمی بسیار نزدیک به نمودار دو جمله ای است و فقط در بازه $\frac{1}{100}<p<\frac{2}{100}$ آن هم به اندازه 0.001 نمودار پوآسون بیشتر است. اما نمودار توزیع نرمال اصلا نزدیک نیست و خیلی پایین تر از دو نمودار دیگر است و دقیق نیست. آن هم به خاطر مقدار p که 0.008 هست و خیلی نزدیک به 0 است و لذا توزیع پوآسون تقریب بهتری خواهد بود.

توجه شود که این مقایسه ها اطراف نقطه $p=\frac{2}{100}$ بررسی شده و در جاهای دیگر هر سه نمودار تقریبا صفر اند.

③ ① برای اینکه نمره دانشجویی جزو 10٪ بالا باشد، یعنی باید از 90٪ دانشجویان نمره‌اش بیشتر باشد، پس باید حساب کنیم که 90٪ ابتدایی دانشجویان حداکثر چه نمره ای دارند تا دانشجوی دیگر از آن بالاتر باشد و لذا جزء 10٪ بشود. از طرفی می‌دانیم تمام نمرات حداقل برابر 0 می‌باشند.

پس نمره حداکثر را برای 90٪ دانشجویان x می‌گیریم و داریم:

$$P\{0<X<x\}=1-\frac{10}{100}=0{,}9 \xrightarrow{\text{نرمال}} F_X(x)-\cancel{F_X(0)}^{\,0}=0{,}9$$

پس باید مقداری را پیدا کنیم که CDF آن در توزیع نرمال با میانگین 80، انحراف معیار 12 برابر 0,9 شود، یعنی:

$$x_{0,9}=F^{-1}(0{,}9)$$

یعنی ما صدک $100\times0{,}9=90$ ام این توزیع را می‌خواهیم. برای یافتن صدک از تابع scipy.stats.norm.ppf استفاده می‌کنیم.

② چارک دوم و سوم به ترتیب یعنی صدک 50 ام و 75 ام. پس بازه از نمرات را می‌خواهیم که احتمال رخ دادن آن‌ها میان $\frac{50}{100}=0{,}5$ ، $\frac{75}{100}=0{,}75$ باشد، در نتیجه به همان تابع ppf بالا مقادیر 0,5 و 0,75 را می‌دهیم تا صدک ها را به ما بدهند و بازه نمرات برابر می‌شود با: (صدک 75 ام ، صدک 50 ام)

③ در توزیع پیوسته‌ی نرمال، طبق فرمول، احتمال اینکه نمره یعنی X میان مثلا 80 و 90 باشد برابر است با:

$$P\{80<X<90\}=F_X(90)-F_X(80)$$

پس ما به تابع scipy.stats.norm.cdf نیاز خواهیم داشت تا CDF را برای این 2 مقدار 80 و 90 پیدا کنیم و احتمال مطلوب را محاسبه کنیم.

---

④ نحوه کد دقیقا مانند سوال 2 است با تنها این تفاوت که در اینجا $n=7072$ و $P=0{,}45$ است و لذا مقادیر λ، μ، σ^2 نیز فرق خواهند کرد.

با توجه به توضیحات سوال 2، در اینجا که $P=0{,}45$ است و خیلی به 0,5 نزدیک است، همینطور با توجه به نمودارها، توزیع نرمال تقریب خیلی بهتر و دقیق‌تری برای دوجمله‌ای خواهد بود.

طبق نمودار توزیع ها نیز، اطراف نقطه $7072\times0{,}45=3182{,}4$ که بررسی کنیم (بقیه جاها نمودار تقریبا صفر است)، نمودار نرمال خیلی نزدیک‌تر به دوجمله‌ای است و نمودار پوآسون اصلا نزدیک نیز نیست و خیلی پایین‌تر از دو نمودار است. البته باید توجه داشت که در این سوال ما 0,05 تا 0,5 فاصله داریم و خطا و فاصله‌ی تقریب و نمودار نیز مشهود و خیلی بیشتری نسبت به سوال 2 دارد که آنجا ما 0,008 تا 0 فاصله داشتیم و تقریب ما نزدیک‌تر بود.