دانشگاه تهران، دانشکده مهندسی برق و کامپیوتر

آمار و احتمال مهندسی

تمرین کامپیوتری اول ــ موضوع تمرین

طراح: علی عطاءاللهی

سوپروایزر: علی محمدی

تاریخ تحویل: ۲۵ آبان ۱۴۰۲

۱. رابطه‌ی توزیع دوجمله‌ای با برنولی (۴۰) نمره

مقدمه

می‌دانیم توزیع دوجمله‌ای، معادل مجموع نتایج حاصل از تکرار نمونه‌برداری از یک توزیع برنولی است. همچنین می‌دانیم اگر X یک متغیر تصادفی از توزیع برنولی با میانگین/احتمال موفقیت p و Y متغیر تصادفی از توزیع دوجمله‌ای، حاصل جمع n بار تکرار نمونه‌برداری از توزیع X باشد، آنگاه میانگین و واریانس توزیع Y به شکل زیر قابل محاسبه است:

$$E[Y] = np$$

$$Var(Y) = np(۱ - p)$$

در بخش اول، قصد داریم این روابط را به صورت عملی بررسی کنیم. ابتدا تابعی برای نمونه برداری از توزیع دوجمله‌ای با استفاده از توزیع برنولی می‌نویسیم. سپس به ازای مقادیر مختلف p به عنوان میانگین توزیع برنولی مورد استفاده، تعداد زیادی نمونه از توزیع دوجمله‌ای برداشته و سپس میانگین و واریانس را محاسبه می‌کنیم. با مقایسه‌ی نمودارهای میانگین و واریانس به ازای مقادیر مختلف p با نمودار روابط داده شده برای میانگین و واریانس توزیع دوجمله‌ای، صدق این روابط را نشان خواهیم داد.

سوال

فرض کنید قصد داریم m نمونه از توزیع دوجمله‌ای متناظر با n بار تکرار یک توزیع برنولی تولید کنیم. برای این کار نیاز داریم m دسته n تایی از توزیع برنولی مذکور نمونه برداری کرده و مجموع نمونه‌های هر دسته را محاسبه و به عنوان مقدار خروجی بازگردانیم.

۱ــ تابع نمونه‌برداری از توزیع دوجمله‌ای را بر مبنای توزیع برنولی پیاده‌سازی کنید.

راهنمایی: برای ایجاد این دسته بندی، با هدف پرهیز از استفاده از حلقه for، در ابتدا $n \times m$ نمونه از توزیع برنولی گرفته و با یک تغییر ابعاد، آن‌ها را به یک ماتریس m در n تبدیل و مجموع هر سطر را محاسبه می‌کنیم.

توجه داشته باشید که این تابع همانند $numpy.random.binomial$ عمل خواهد کرد.

توابع مورد نیاز: $numpy.random.choice, numpy.array, numpy.sum$

۲ــ به ازای مقدار ثابت $n = ۵۰۰$، به ازای هر p (از صفر تا صد)، ۵۰۰۰ نمونه از توزیع دوجمله‌ای برداشته و میانگین و واریانس تئوری (با استفاده از فرمول) و عملی (با استفاده از تابع ساخته شده در بالا) را محاسبه کنید.

۳ــ با استفاده از $matplotlib.pyplot$ نمودار مربوط به مقادیر تئوری و عملی میانگین و واریانس را رسم کنید .

۴ــ با توجه به موارد بالا مقادیر تئوری و عملی میانگین و واریانس را مقایسه کنید.

۲. تخمین توزیع دوجمله‌ای به کمک توزیع‌های نرمال و پواسون (۲۰) نمره

مقدمه

می‌دانیم توزیع دوجمله‌ای تحت شرایط ویژه‌ای به صورت حدی به توزیع‌های پواسن و نرمال میل می‌کند. بنابراین جهت محاسبه‌ی احتمال توزیع دوجمله‌ای در مسائل با ابعاد بزرگ می‌توان از تقریب‌های پواسن و نرمال استفاده نمود. همچنین می‌دانیم که تقریب توزیع دوجمله‌ای با استفاده توزیع نرمال برای احتمال‌های نزدیک به ۰/۵ مناسب است و برای احتمال‌های نزدیک به صفر تقریب پواسن مناسب‌تر است.

در این بخش انتظار داریم تقریب پواسن از تقریب نرمال عملکرد بهتری داشته باشد. همچنین قصد داریم در قالب یک مثال واقعی، درستی و شرایط استفاده هر یک از تقریب‌های مذکور را به دقت بررسی کنیم.

سوال

مطابق گزارش آمار تصادف و حوادث جاده‌ای پژوهشکده‌ی مرکز آمار ایران، روزانه در شهر تهران به صورت میانگین ۲۵۰ تصادف رخ می‌دهد. مطابق مطالعات صورت گرفته مرگ یک فرد حاضر در تصادف یک متغیر تصادفی است که از توزیع برنولی با احتمال ۰/۰۰۸ پیروی می‌نماید. مطابق مطالب مطرح شده در بخش رابطه توزیع دوجمله‌ای با برنولی می‌دانیم که می‌توان تعداد مرگ و میرهای روزانه ناشی از تصادف شهروندان تهرانی را یک متغیر تصادفی دوجمله‌ای به صورت زیر تعریف کرد:

$$X \sim Bin(250, 0.008)$$

که در رابطه‌ی مذکور مقصود از X همان تعداد مرگ و میر شهروندان تهرانی در یک روز بر اثر تصادف است.

همانطور که پیش از این اشاره شد، با توجه به احتمال اندک مرگ و میر و تعداد زیاد تصادفات روزانه در شهر تهران، متغیر تصادفی X را می‌توان با استفاده از توزیع پواسن تقریب زد:

$$\hat{X}_1 \sim Poi(\mu); \;\; \mu = 250 \times 0.008 = 2$$

از طرفی می‌دانیم با توجه به تعداد زیاد تصادفات روزانه در تهران، می‌توان متغیر تصادفی X را با استفاده از توزیع نرمال نیز تقریب زد:

$$\hat{X}_2 \sim \mathcal{N}(\mu, \sigma^2); \;\; \mu = 250 \times 0.008 = 2; \;\; \sigma^2 = 250 \times 0.008 \times (1 - 0.008) = 1.984$$

۱ ــ .با استفاده از شبیه‌سازی توزیع‌ها (دوجمله‌ای، پوآسون، نرمال) درستی ادعاهای مذکور را نشان دهید (نمودارهای مربوط به آنها را رسم کنید)

راهنمایی: با استفاده از توابع پیشفرض به رسم توزیع‌ها پرداخته و در انتها با استفاده از تحلیل‌های کیفی به مقایسه‌ی عملکرد هر تقریب می‌پردازیم. همچنین با توجه به تعداد تصادفات رخ داد روزانه تهران (۲۵۰) و احتمال رخ دادن مرگ و میر در هر تصادف (۰۰۸.۰) منطقی است که نمودار توزیع رسم شده در محدوده‌ی $2 = 0.008 \times 250$ رسم گردد. زیرا با توجه به ماهیت توزیع‌ها در خارج از محدوده‌ی ذکر شده احتمال تقریبا صفر خواهد بود.

توابع مورد نیاز: $numpy.random.binomial, poisson.pmf$

۲ ــ کدام تقریب سازگارتر است؟ استدلال خود را مطرح نمایید.

## ۳. اهمیت توزیع نرمال (بخش اول)	(۴۰) نمره

مقدمه

در ابتدا بررسی می‌کنیم که اهمیت و علت شهرت توزیع نرمال چیست. علت شهرت توزیع نرمال این است که اگر تعداد زیادی متغیر تصادفی مستقل را با هم جمع کنیم توزیع حاصل قطعا نرمال خواهد بود. صحت گزاره‌ی مذکور از کلمه‌ی ”زیاد” سرچشمه می‌گیرد. چه تعداد متغیر تصادفی ”زیاد” است؟ در واقع گزاره‌ی دقیق از نظر ریاضی این است که توزیع حاصل از جمع متغیرهای تصادفی مستقل وقتی تعدادشان به بی‌نهایت میل کند، به توزیع نرمال میل می‌کند. اما در واقعیت اگر توزیع متغیرها خیلی متمایل به یک سمت بازه نباشد (برخلاف توزیع درآمد که افراد کمی درآمد بسیار بسیار زیادی دارند) جمع تعداد کمی متغیر تصادفی توزیعی بسیار نزدیک به توزیع نرمال دارد. اسم قضیه‌ی گفته‌شده، قضیه‌ی ”حد مرکزی” است.

به همین دلیل است که توزیع نرمال تا این حد برای مطالعات آماری نقش ”مرکزی” دارد. برای مثال می‌توان تصور کرد که نمره‌ی دانشجویان در کلاس احتمال یک متغیر تصادفی است که خود از جمع موارد بسیار زیادی تشکیل شده است برای مثال علاقه‌ی دانشجویان به ریاضی، پیشینه‌ی تحصیلی آن‌ها، منابع مختلفی که برای مطالعه انتخاب می‌کنند و... که هر یک از این متغیرها ممکن است نرمال نباشد (در واقع خیلی از آن‌ها نمی‌توانند نرمال باشند چون گسسته هستند) اما جمع همه‌ی آن‌ها که نمره‌ی نهایی را می‌سازد توزیع نرمال دارد.

سوال

فرض کنید نمره‌های امتحان آمار و احتمال مهندسی دانشگاه تهران در سال ۱۴۰۲ توزیع نرمال با میانگین ۸۰ و انحراف معیار ۱۲ دارند. همچنین تعدادی از نمرات به دلیل نمره‌ی امتیازی بالای ۱۰۰ هستند (فرض کنید تعداد دانشجویان خیلی زیاد است).

۱ــ اگر یک دانشجو جزو ده درصد بالای کلاس باشد، نمره‌ی او حداقل چقدر است؟

۲ــ چه بازه‌ای از نمرات در چارک دوم و سوم قرار دارد؟

۳ــ احتمال این که نمره بین ۸۰ و ۹۰ باشد چقدر است؟

۴ــ فرض کنید نمرات امتحان فیزیک ۲ در سال ۱۴۰۲ دارای توزیع یکنواخت باشد. همچنین نمرات امتحان برنامه‌نویسی پیشرفته و گسسته در همین سال به ترتیب دارای توزیع نمایی و پوآسون باشد. حال از نمرات این سه درس (که مستقل از هم هستند) تعداد زیادی نمونه بگیرید و سپس هیستوگرام حاصل جمع را رسم کنید و نشان دهید که هیستوگرام به سمت توزیع نرمال حرکت می‌کند. دقت کنید تمام مقادیر و پارامترهای مربوط به توزیع‌ها را به طور دلخواه در نظر بگیرید به گونه‌ای که قابل توجیح باشند. همچنین تمام فرضیات خود را در گزارش ذکر کنید (امتیازی).

کتابخانه‌های مورد نیاز: $scipy.stats, numpy$

بخش امتیازی: $numpy.random.normal, numpy.random.exponential, numpy.random.poisson$

## ۴. اهمیت توزیع نرمال (بخش دوم)	(۲۰) نمره

مطابق گزارش آمار تصادف و حوادث جاده‌ای پژوهشکده‌ی مرکز آمار ایران، روزانه در شهر تهران به صورت میانگین مطابق گزارش مرکز آمار ایران، زخمی شدن یک شهروند تهرانی در تصادفات درون شهری از یک توزیع برنولی با احتمال $p = ۰/۴۵$ پیروی می‌نماید و همچنین ماهانه به صورت میانگین ۷۰۷۲ تصادف در شهر تهران رخ می‌دهد. اگر تعداد افراد زخمی شده ناشی از تصادفات درون شهری تهران را متغیر تصادفی Y با توزیع دوجمله‌ای در نظر بگیریم مشابه قبل داریم:

$$Y \sim Bin(۷۰۷۲, ۰/۴۵)$$

مشابه قبل می‌دانیم تقریب‌های پواسن و نرمال از متغیر تصادفی Y به شرح زیر می‌باشد:

$$\hat{Y}_1 \sim Poi(3182/59)$$

$$\hat{Y}_2 \sim \mathcal{N}(3182/59, 41/84)$$

همچنین می‌دانیم در توزیع‌های دوجمله‌ای، برای احتمال‌های نزدیک به ۰/۵ تقریب نرمال مناسبی است اما تقریب‌های پواسن برای احتمال‌های نزدیک به ۰ دارای اعتبار می‌باشند. بنابراین انتظار داریم تقریب نرمال برای مسئله‌ی حاضر عملکرد بهتری داشته باشد.

۱ ــ با استفاده از شبیه‌سازی توزیع‌ها (دوجمله‌ای، پوآسون، نرمال) درستی ادعاهای مذکور را نشان دهید. (نمودارهای مربوط به آنها را رسم کنید)

توابع مورد نیاز: $numpy.random.binomial, poisson.pmf, scipy.stats.poisson.pmf,$

$scipy.stats.norm.pdf$

۲ ــ کدام تقریب سازگارتر است؟ استدلال خود را مطرح نمایید.

نحوه‌ی تحویل

فایل یا فایل‌های py. یا ipython. حاوی کدها و فایل PDF گزارش را در یک فایل زیپ با نام CA#1-STDNo.zip قرار داده و روی سایت درس بارگزاری کنید. دقت کنید که مراحل کدزنی و نتایج شما باید به طور کامل در گزارش بیان شده باشد.

.

موفق باشید