گزارش تمرین کامپیوتری ۱ درس آمار و احتمال

پریسا یحیی‌پور فتیده　　　شماره دانشجویی: ۸۱۰۱۰۱۵۵۱　　　استاد توسلی‌پور

سوال ۱:

تابع generate_binomial تابعی است که تعداد نمونه های هر توزیع دو جمله ای و تعداد کل توزیع های دوجمله ای و احتمال موفقیت را ورودی می گیرد سپس یک آرایه نامپای یک بعدی به عنوان خروجی ریترن می کند که مجموع نمونه های هر توزیع احتمال است. توجه شود که برای جلوگیری از استفاده کردن از حلقه for به تعداد n*m نمونه تصادفی می گیریم و سپس آرایه نامپای یک بعدی را به آرایه دوبعدی یا ماتریس تبدیل میکنیم و جمع هر سطر نشان دهنده توزیع دوجمله ای است.

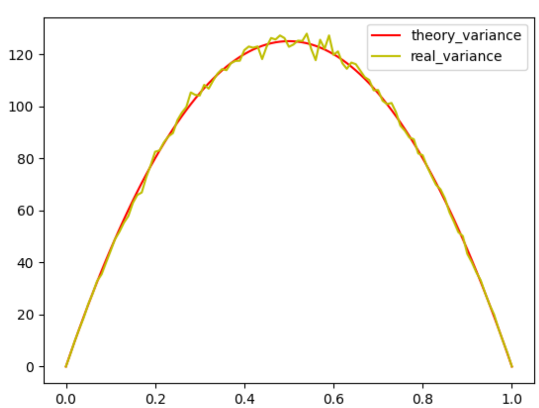
به کمک یک حلقه فور p های مطلوب را تولید کرده تا به عنوان ورودی به تابع بدهیم و خروجی های مختلف بگیریم و داده های مورد نیاز را محاسبه می کنیم.

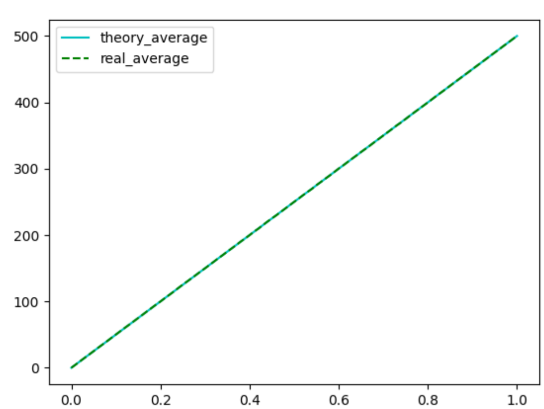
میانگین تئوری به کمک فرمول np محاسبه می شود در حالی که میانگین عملی را به کمک تابع numpy.mean محاسبه می کنیم که همان فرمول معمول جمع تقسیم بر تعداد است.

واریانس تئوری به کمک فرمول npq محاسبه می شود در حالی که واریانس عملی به کمک تابع numpy.var محاسبه می شود که از فرمول زیر پیروی میکند:

$$\sigma^2 = \frac{\Sigma\,(x-\mu)^2}{N}$$

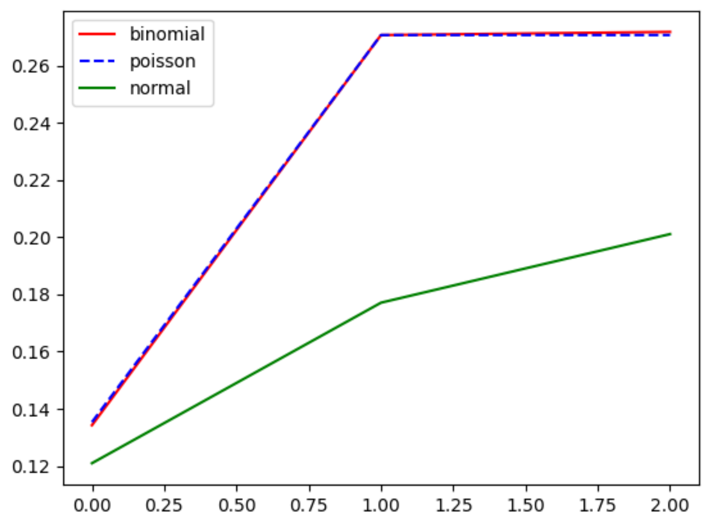
سپس نمودار میانگین و واریانس را رسم می کنیم. میانگین به هردو روش محاسبه یکسان و برهم منطبق است اما واریانس در بیشتر جاها منطبق است برای احتمال موفقیت اطراف ۰.۵ واریانس تئوری و عملی باهم متفاوت هستند به این دلیل که توزیع دوجمله ای برای این احتمال موفقیت برای تعداد های بالا به توزیع نرمال میل می کند بنابراین تفاوت دارند.





سوال ۲:

ابتدا مقادیر مورد نیاز را تولید میکنیم که در اینجا تا ۲ محاسبه شده است سپس به کمک توابع آماده pmf احتمال رخ دادن هر حالت را در توزیع های مختلف محاسبه و در نهایت رسم و با هم مقایسه می کنیم.



همان طور که از نمودار مشخص است توزیع دوجمله ای برای احتمال کم که تقریبا نزدیک به صفر است به خوبی با توزیع پواسون قابل اندازه گیری است. در حالی که توزیع نرمال تقریب خوبی برای توزیع دوجمله ای نیست و این موضوع طبق نمودار قابل استنباط است.

سوال ۳:

بخش اول: برای محاسبه این نمره باید احتمال را برابر ۰.۱ بگذاریم از انجایی که نمیتوانیم احتمال متغیر های تصادفی بزرگتر از یک مقدار را محاسبه کنیم بنابراین از cdf استفاده میکنیم که قابل محاسبه است سپس باید توزیع نرمال استاندارد به کمک جدول محاسبه کنیم که به ازای چه مقداری خروجی ان ۰.۹ است. سپس معکوس عملیاتی که برای توزیع نرمال استاندارد انجام دادیم را محاسبه میکنیم تا به نمره برسیم که ۹۵.۶ است.

بخش دوم: تعریف چارک های در عکس زیر نوشته شده است. مشابه قسمت قبلی به کمک توزیع نرمال استاندارد و جدول cdf آن چارک ها را محاسبه میکنیم.

بخش سوم: همان طور که میدانیم برای پیدا کردن احتمال رخ دادن متغیر تصادفی در یک بازه به علت پیوسته بودن توزیع از انتگرال تابع cdf استفاده میکنیم.

راه حل تشریحی همه بخش ها در تصویر زیر آمده است.

$P\{X > x_0\} = \frac{1}{10}$ ①

$$P\{X > x_0\} = 1 - \underbrace{P\{X < x_0\}}_{\text{CDF تعریف}} = 1 - F_X(x_0) = \frac{1}{10} \Rightarrow F_X(x) = \frac{9}{10}$$

$$\Rightarrow \phi_0(Z) = 0.7 \longrightarrow Z = 1.3 = \frac{x_0 - 80}{12} \Rightarrow x_0 = 95.6$$

②

چارک دوم ← صدک 50 ام ← میانگین ← 80

چارک سوم ← صدک 75 ام ←

$$F^{-1}(0.75) = ?$$

$$\phi(x) = 0.75 \longrightarrow x = 0.7 \Rightarrow \frac{x - 80}{12} = 0.7 \Rightarrow x = 88.4$$

بازه ⇐ $[80, 88.4]$

③

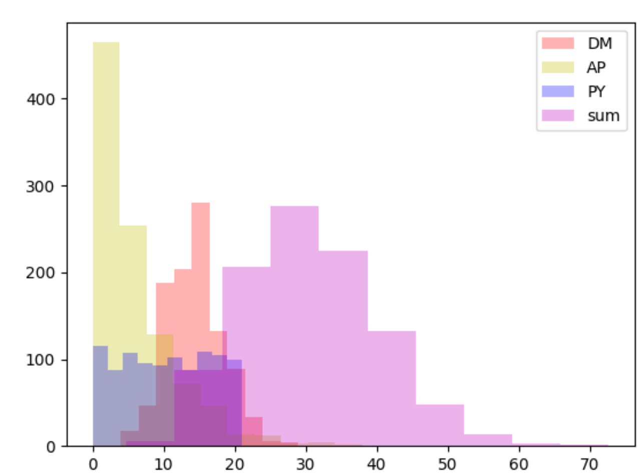
$$P\{80 \leqslant X \leqslant 90\} = F_X(90) - F_X(80) = \int_{80}^{90} \frac{1}{12\sqrt{2\pi}} e^{\frac{-(x-80)^2}{288}} dx = 0.297$$

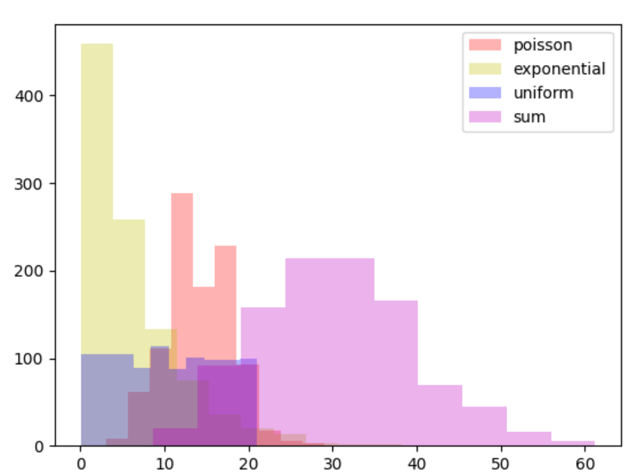
بخش چهارم: برای توزیع پواسون از پارامتر ۱۴ به طور حدسی، برای توزیع یکنواخت از ۰ و ۲۱ استفاده کردیم چون نمره های بین ۰ تا ۲۰ می توانند باشد و تابع بازه باز است پس ۲۱ می زنیم. برای توزیع نمایی از پارامتر ۶ استفاده کردیم. دلیل کمتر بودن از توزیع پواسون این است که پارامتر ورودی برای توزیع نمایی معکوس است.

سپس از همه توزیع ها به تعداد ۱۰۰۰ نمونه تصادفی بر میداریم و نمودار هر کدام را می کشیم سپس این داده های تصادفی را با هم جمع کرده تا با توزیع نرمال مقایسه کنیم.

تابع hist آرایه نامپای ورودی را به گونه ای به نمودار تبدیل می کند که محور x نشان دهنده داده های متمایز و محور y نشان دهنده فراوانی داده های متمایز باشد.

همان طور که از نمودار هم پیداست نمودار جمع این نمرات مانند توزیع نرمال است.





نکته ای که باید به آن توجه کنیم این است که با هربار ران کردن کد به خروجی خای متفاوت می رسیم چون نمونه های تصادفی تغییر می کنند.

سوال ۴:

این سوال مشابه سوال ۲ با پارامتر های متفاوت است. همان طور که در سوال دو دیدیم توزیع دوجمله ای با احتمال موفقیت کم اما تعداد بالا قابل تقریب زدن با توزیع پواسون است اما در اینجا از انجایی که احتمال موفقیت نزدیک به ۰.۵ است توزیع پواسون تقریب خوبی برای توزیع دو جمله ای نیست که طبق نمودار این ادعا قابل برداشت است.