



به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده مهندسی برق و کامپیوتر

درس آمار و احتمال مهندسی

استاد: دکتر ربیعی

## تمرین کامپیوتری شماره 1

نام و نام خانوادگی: محمدمهدی عبدالحسینی

شماره دانشجویی: 810198434

آبان ماه 1399

## چکیده:

هدف از این تمرین آشنایی با توزیع های آماری از جمله توزیع نرمال، توابع  $CDF, PDF, PMF$  و تحلیل داده، در زبان برنامه نویسی پایتون می باشد. که جهت بررسی موضوع در بخش اول ابتدا به توزیع نرمال و تحلیل آن پرداخته می شود و سپس در بخش دوم به مقدمه ای بر تحلیل داده و ارائه ی مثال هایی به کاربرد زبان برنامه نویسی پایتون در تحلیل داده ها اشاره میگردد.

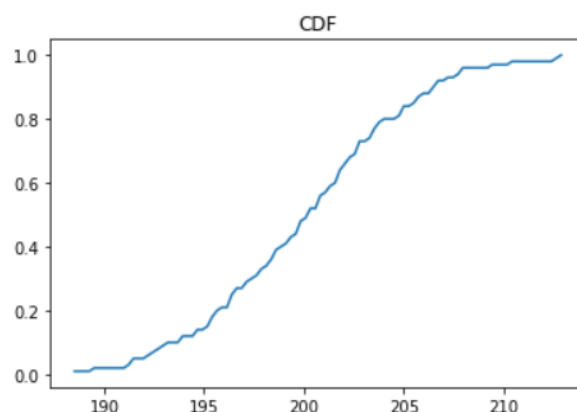
## بخش 1

1: در سوال اول ابتدا یک دنباله تصادفی  $X=[X_1..X_n]$  به ازای  $n=100$  با توزیع نرمال  $X_i \sim N(\mu=200, \sigma^2=25)$  ایجاد میکنیم. سپس با استفاده از هیستوگرام این نمونه ها، CDF آن ها رسم شده است. برای انجام این کار میتوان از تابع `np.random.normal()` استفاده کرد. بوسیله تابع `np.histogram()` دو خروجی لازم برای رسم CDF بدست میاید. این دو خروجی در واقع طول بازه ها و تعداد متغیرهای تصادفی در آن بازه ها میباشد.

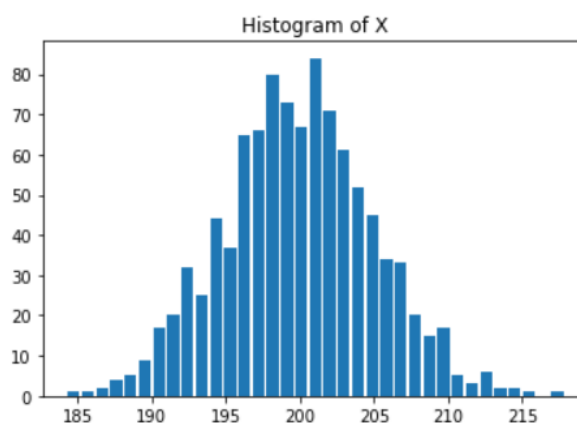
اکنون برای تعیین CDF میتوان بصورت زیر عمل کرد:

تعداد مقادیر را در بازه ی ابتدایی تقسیم بر تعداد کل داده ها ( $n$ ) میکنیم و آن را برابر با CDF در این بازه قرار میدهیم. برای بازه ی بعدی همین کار تکرار میشود با این تفاوت که CDF در این بازه با مقدار های قبلی جمع میشود. و همین کار را تا آخرین بازه انجام میدهیم.

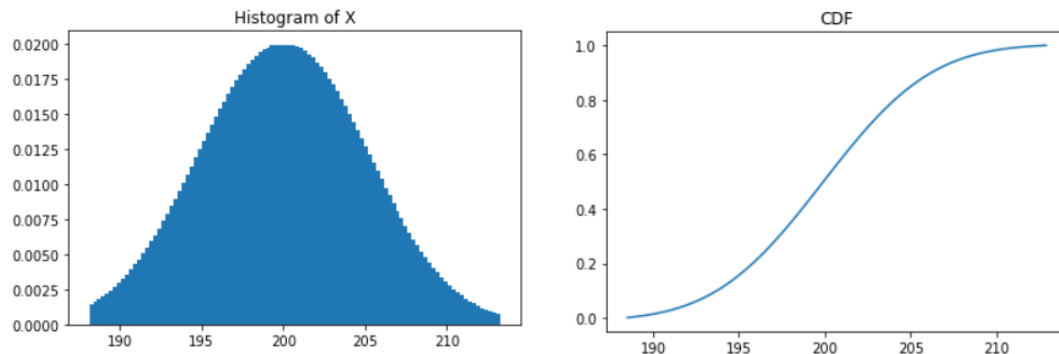
در نهایت نمودار CDF بصورت زیر خواهد بود: (تعداد بازه ها برای رسم CDF ، 100 میباشد).



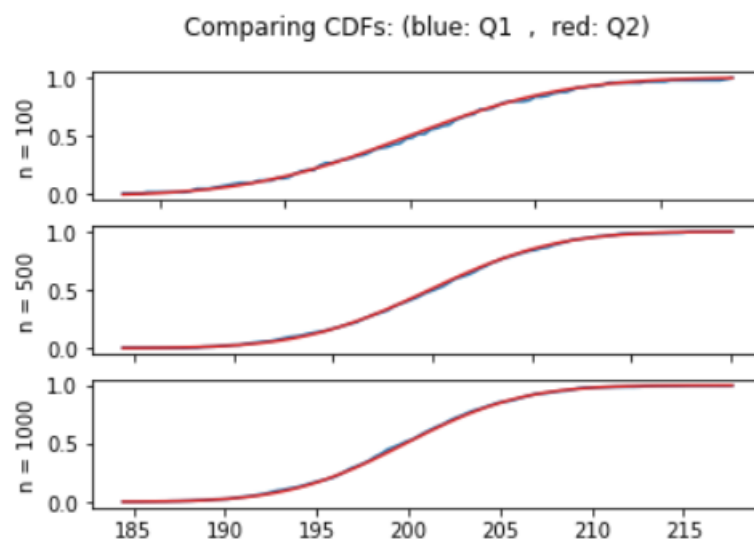
همچنین نمودار هیستوگرام به ازای  $n = 1000$  و طول بازه 35 بصورت زیر است:



2: با استفاده از رابطه (1) که در متن پروژه آمده است، نمودار CDF به بسیار دقیقی بدست میاید و همچنین نمودار هیستوگرام آن به واقعیت نزدیکتر است.



3: در این قسمت با رسم نمودارهای CDF در سه حالت  $n = 100, 500, 1000$  و مقایسه آن با قسمت اول به نتایج جالبی میرسیم. ابتدا نگاهی به نمودارها بیندازیم تا سپس با توجه به آنها مسئله را تحلیل کنیم.



نمودارهای قرمز رنگ مربوط به نتایج قسمت اول و نمودارهای آبی رنگ مربوط به نتایج قسمت دوم میباشد. با توجه به این که میدانیم نتایج قسمت دوم دقیق تر است اما میتوان با افزایش مقدار  $n$  میتوان خطای قسمت اول را کاهش داد. در واقع اگر  $n$  را بینهایت فرض کنیم، دیگر اثری از نمودار آبی نخواهیم دید و این نشان دهنده این است که خطا صفر میشود. نتیجه دیگری که میتوان از نمودارهای بالا برداشت کرد این است که در صورت افزایش  $n$ ، نمودار واقعی تر میشود. این یعنی متغیرهای بیشتری در حدود میانگین قرار میگیرند و شیب تغییرات افزایش میابد و نمودار در مقادیر کوچکتری به 0 و 1 میرسد.

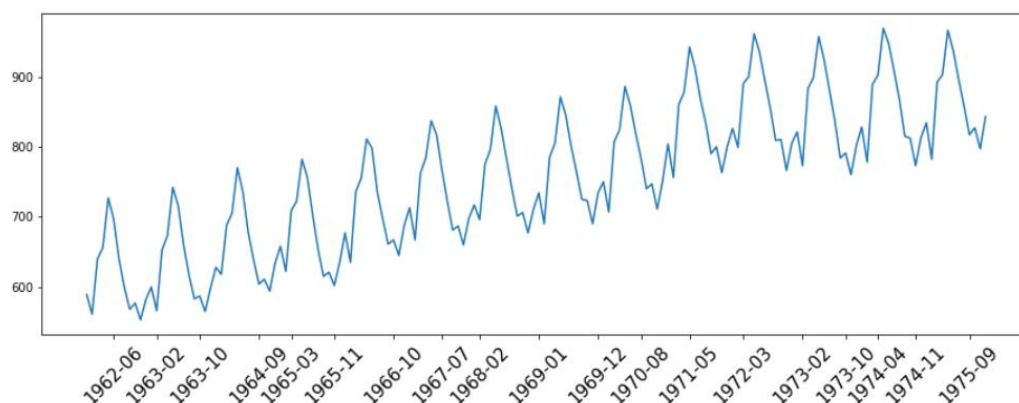
4: با استفاده از نحوه ی بدست آوردن CDF در قسمت 1 ، اگر در حین محاسبه، مقادیری از CDF را که در بازه  $(\mu - \sigma, \mu + \sigma)$  قرار دارد را با هم جمع کنیم و سپس تقسیم بر تعدادشون کنیم، احتمال آنکه میزان بارش بین  $(\mu - \sigma, \mu + \sigma)$  باشد، بدست میاید.

نکته قابل توجه این است که با تغییر مقادیر  $\mu$  و  $\sigma$  به هر نحوی، این احتمال ثابت میماند و تنها با تغییر مقدار  $n$  تغییر میکند. علت کاملاً مشخص است. در واقع این مسئله یک ویژگی ذاتی برای متغیر تصادفی نرمال یا گوسی است و به مقادیر  $\mu$  و  $\sigma$  وابسته نیست. به بیان دیگر با تغییر  $\mu$  ، نمودار تنها در راستای افقی جابجا میشود و تغییری در ظاهر نمودار ایجاد نمیشود. پس احتمال خواسته شده بدون تغییر میماند. در صورت تغییر  $\sigma$  ، پراکندگی داده ها در حول میانگین تغییر میکند. اما چون بازه خواسته شده برای محاسبه احتمال نیز متناسب با آن تغییر میکند، پس مجموع احتمال ها ثابت میماند.

## بخش 2

1: در این قسمت با بررسی فایل tesla-stock-price.csv تعداد دفعاتی که ارزش سهام شرکت در شروع بازه ی خرید و فروش بیشتر از ورودی داده شده باشد بر روی خروجی چاپ میشود. بطور مثال برای ورودی 300، خروجی مقدار 313 خواهد بود.

2: نمودار میزان تولید شیر مزرعه را بر حسب زمان به صورت زیر خواهد بود:



از نمودار بالا میتوان نتیجه گرفت که میزان تولید شیر مزرعه به صورت تناوبی میباشد.

3: نمودار میله ای 10 کشور اول با بیشترین مهاجر به صورت زیر خواهد بود.

