


شرکت‌های تولیدکننده مطرح حوزه تکنولوژی، قبل از معرفی تولیدات جدید خود، معمولاً برای تبلیغ هر چه بیشتر محصول، رویدادی با حضور افراد تأثیرگذار و مطرح دنیای تکنولوژی برگزار می‌کنند. حال، فرض کنید شما در گروه آنالیز داده یکی از این شرکت‌ها فعالیت می‌کنید. برای مراسم معرفی محصول جدید، گروه ارتباطات شرکت، تعداد ۱۰۰۰ ایمیل را به مبلغان حوزه تکنولوژی در سراسر دنیا ارسال می‌کند. بر اساس داده پیشین، مخاطبان این جنس ایمیل‌ها، به حدود ۴۵ درصد از ایمیل‌ها پاسخ مثبت می‌دهند و در مراسم معرفی حضور پیدا می‌کنند. شرکتی که شما در آن مشغول به کار هستید، دچار بودجه‌ای محدود است و برای برگزاری مراسم، قادر است حداکثر سالنی با ظرفیت ۴۳۰ نفر تهیه کند. حال، از شما خواسته شده است احتمال این که حداکثر ۴۳۰ نفر از این ۱۰۰۰ نفر، به دعوت شرکت جواب مثبت بدهند را محاسبه کنید تا شرکت بر اساس آن تصمیم‌گیری کند.

۱- با استفاده از تابع pmf توزیع دوجمله‌ای که در سوال ۱، قسمت ۵، پیاده‌سازی کردیم  تابع محاسبه CDF این توزیع را با استفاده از توابع پایه در پایتون، پیاده‌سازی کنید و احتمال خواسته‌شده را گزارش کنید.

به علت پیچیدگی محاسبات و محدود بودن منابع سخت‌افزاری، شرکت از شما می‌خواهد به جای استفاده از توزیع دوجمله‌ای، از توزیع نرمال برای تقریب مقدار احتمال خواسته‌شده استفاده کنید.

۲- با استفاده از CDF توزیع نرمال و بدون اعمال تصحیح پیوستگی، احتمال خواسته‌شده را تقریب بزنید و آن را به همراه مقدار خطا نسبت به احتمال حاصل‌شده در قسمت قبل گزارش کنید.

۳- بار دیگر با استفاده از CDF توزیع نرمال و با اعمال تصحیح پیوستگی، احتمال خواسته‌شده را تقریب بزنید و آن را به همراه مقدار خطا گزارش کنید.

۴- در این قسمت، ابتدا مقدار CDF برای مقادیر  $X = 0, 1, 2, 3, \dots, 1000$  را با هر سه روش بالا به دست‌آورد و مقادیر خطا را یکبار بین مقادیر CDF توزیع دوجمله‌ای و توزیع نرمال بدون تصحیح پیوستگی و یکبار بین مقادیر CDF توزیع دوجمله‌ای و توزیع نرمال با تصحیح پیوستگی به دست‌آورد. دو دسته خطای به‌دست‌آمده را بر حسب مقدار متغیر تصادفی در یک شکل رسم کنید. دو نمودار حاصل را تحلیل کنید. ( $X$  همان متغیر تصادفی است.)

۵- با استفاده از مقادیر CDF محاسبه‌شده، نمودار CDF بر حسب مقدار متغیر تصادفی را برای مقادیر بازه  $X \in [440, 460]$  برای هر سه روش مذکور، در یک شکل رسم کنید و نمودارها را مقایسه کنید.

۶- برای مقادیر  $X = 0, 1, 2, 3, \dots, 1000$  بار دیگر مقدار CDF را با دو روش توزیع دوجمله‌ای و توزیع نرمال با تصحیح پیوستگی محاسبه کنید و زمان سپری‌شده برای محاسبه هر یک از مقادیر CDF را برای هر دو روش ذخیره کنید و بر حسب مقدار متغیر تصادفی در یک شکل رسم کنید. دو نمودار به دست‌آمده را تحلیل کنید.

راهنمایی: برای محاسبه زمان سپری‌شده یک عملیات با دقت بالا، از فرمت زیر استفاده کنید:

```
import time
def get_elapsed_time(*args):
    start = time.perf_counter()
    # your function or operation
    end = time.perf_counter()
    return end - start
```

با توجه به محدود بودن منابع سخت‌افزای شرکت، از شما خواسته شده‌است یک استراتژی تعیین کنید که مشخص کند برای محاسبه CDF چه مقداری از  $X$  در بازه  $[400, 600]$  از توزیع دوجمله‌ای و برای چه مقداری، از توزیع نرمال با تصحیح پیوستگی استفاده کنیم. بدین منظور، فرض کنید هزینه پرداختی برای هر ثانیه افزایش زمان محاسبات برابر با ۱۰۰ واحد و هزینه پرداختی برای هر واحد خطا در محاسبه CDF برابر با ۱۰۶ واحد باشد.

۷- ابتدا برای استفاده از مقادیر به دست‌آمده در قسمت‌های ۴ و ۶، مقادیر افزایش زمان محاسبات هنگام استفاده از توزیع دوجمله‌ای نسبت به توزیع نرمال با تصحیح پیوستگی و خطای محاسبه CDF هنگام استفاده از توزیع نرمال با تصحیح پیوستگی را برای مقادیر بازه  $X \in [400, 600]$  به دست‌آورد. سپس تابعی پیاده‌سازی کنید که با استفاده از هزینه‌های فرض‌شده برای هر یک از دو مورد، اولین مقدار  $X$  را بازگرداند (با شروع از  $X = 400$  و در بازه  $X = [400, 600]$ ) که محاسبه CDF آن با استفاده از توزیع دوجمله‌ای مقرون به صرفه نیست. نقطه به دست‌آمده را گزارش کنید.

راهنمایی: محاسبه CDF با استفاده از توزیع دوجمله‌ای هنگامی مقرون به صرفه نیست که هزینه افزایش زمان محاسبات، از هزینه خطای محاسبه CDF توسط توزیع نرمال با تصحیح پیوستگی، بیشتر شود.

۸- نمودار هزینه نهایی (اختلاف هزینه افزایش محاسبات و هزینه خطای محاسبات) بر حسب مقادیر  $X \in [400, 600]$  را رسم کنید. با توجه به نمودار و نقطه به دست‌آمده، یک استراتژی ساده برای محاسبه CDF برای مقادیر  $X \in [400, 600]$  تعیین کنید و مشخص کنید آیا محاسبه CDF برای نقطه  $X = 430$  توسط توزیع دوجمله‌ای که در قسمت اول انجام شد، مقرون به صرفه بوده‌است یا خیر.