

در بسیاری از انتخابات در سراسر دنیا، پس از اتمام رای‌گیری، فرآیندهای مختلفی برای اطمینان از صحت برگزاری انتخابات انجام می‌شود که در آن‌ها، از بین تمامی حوزه‌های موجود اخذ رای، تعدادی حوزه انتخاب می‌شوند و آرای آن‌ها مورد بررسی قرار می‌گیرند. حال، فرض کنید در انتخابات مورد نظر، در کل $N=100$ حوزه رای‌گیری وجود دارد که در $k=20$ عدد از آن‌ها تقلب رخ داده است و همچنین، نهاد حسابرسی، $m=40$ ناحیه را مورد بررسی قرار می‌دهد.

۱- توزیع فوق هندسی مربوط به فرآیند حسابرسی را با $n=150$ نمونه شبیه‌سازی کنید و نمودار توزیع آن را رسم کنید.

۲- تابعی پیاده‌سازی کنید که به ازای تعداد نمونه $n=100$ تا $n=10000$ با افزایش‌های ۵۰ واحدی، مقادیر تئوری (با استفاده از روابط ریاضی) و عملی (با استفاده از توزیع شبیه‌سازی‌شده) میانگین و واریانس تعداد تقلبی که توسط نهاد حسابرسی یافت می‌شود را محاسبه کند و بازگرداند. (پارامترهای توزیع فوق هندسی همانند مقادیر گفته شده در صورت سوال است).

۳- مقادیر عملی میانگین و واریانس به دست‌آمده برای مقادیر مختلف n را به همراه مقادیر تئوری در یک نمودار رسم کنید و آن‌ها را مقایسه کنید.

۴- به ازای تعداد نمونه ثابت $n=1000$ ، به ازای مقادیر $m=[40, 60, 80, 100]$ ، توزیع‌های فوق هندسی مربوطه را شبیه‌سازی کنید و نمودار آن‌ها را در یک شکل رسم کنید. نمودارهای به دست‌آمده را تحلیل کنید.

در این قسمت، قصد داریم رابطه بین توزیع دوجمله‌ای و فوق هندسی را بررسی کنیم. توزیع فوق هندسی معمولاً برای شبیه‌سازی نمونه‌گیری بدون جایگذاری استفاده می‌شود، در صورتی که در توزیع دوجمله‌ای، نمونه‌گیری با جایگذاری مناسب اتفاق می‌افتد. حال، هنگامی که مقدار جمعیت N در توزیع فوق هندسی بسیار بزرگ شود، اثر عدم جایگذاری بعد از نمونه‌گیری کاهش پیدا می‌کند و با میل کردن N به سمت بی‌نهایت، توزیع فوق هندسی به توزیع دوجمله‌ای میل می‌کند. اثبات تئوری این موضوع در [این لینک](#) قابل مشاهده است. در ادامه، این موضوع را به صورت عملی مشاهده خواهیم کرد.

۵- با استفاده از توابع پایه در پایتون، توابعی مجزا پیاده‌سازی کنید که پارامترهای هر کدام از توزیع‌های فوق هندسی و دوجمله‌ای را به همراه آرایه‌ای از مقادیر متغیر تصادفی، به عنوان ورودی دریافت کند و مقدار pmf توزیع مربوطه را برای مقادیر آن بازگرداند. (مقادیر متغیر تصادفی را از ۰ تا ۲۰ در نظر بگیرید).

راهنمایی: می‌توانید از توابع **factorial** و **comb** از کتابخانه **math** برای پیاده‌سازی توابع کمک بگیرید.

۶- با استفاده از توابع پیاده‌سازی شده، منحنی pmf توزیع فوق هندسی و دوجمله‌ای را به ازای مقادیر $N=[100, 500, 1000, 10000]$ رسم کنید. (N برابر با تعداد حوزه‌های رای‌گیری است. در این بخش، برای هر مقدار N یک شکل جدا در نظر بگیرید و در هر شکل، توزیع دوجمله‌ای و فوق هندسی مربوطه را با هم رسم کنید).

۷- نمودارهای به دست‌آمده توزیع فوق هندسی و دوجمله‌ای را برای مقادیر مختلف N (جمعیت) تحلیل کنید و نتیجه به دست‌آمده را گزارش و با نتیجه تئوری مقایسه کنید.