**📈 PHẦN 3 — PHÂN PHỐI XÁC SUẤT & MÔ PHỎNG THỐNG KÊ**

*(Probability Distributions & Simulation in Python)*

**🎯 1. Thư viện liên quan / Related Libraries**

| **Lệnh / Thư viện** | **Giải thích (Việt)** | **English Explanation** | **Khi nào dùng / When to use** |
| --- | --- | --- | --- |
| import numpy as np | Thư viện mảng số học | Numerical operations and random simulation | Dùng để sinh dữ liệu ngẫu nhiên |
| import matplotlib.pyplot as plt | Thư viện trực quan hóa | Visualization library | Để vẽ phân phối (PDF, histogram) |
| from scipy.stats import \* | Gọi toàn bộ phân phối xác suất | Import all distributions from SciPy | Khi làm việc với PDF, CDF, PPF |

**🧮 2. Các hàm tạo ngẫu nhiên trong numpy.random**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số chính** |
| --- | --- | --- | --- | --- |
| np.random.rand(n) | Sinh n số thực [0,1) | Generate uniform [0,1) | Khi cần số ngẫu nhiên chuẩn hóa | n: số phần tử |
| np.random.randint(low, high, size) | Sinh số nguyên ngẫu nhiên | Random integers | Khi mô phỏng dữ liệu rời rạc (như chiều cao, tuổi) | low, high, size |
| np.random.randn(n) | Sinh số theo phân phối chuẩn (μ=0,σ=1) | Generate standard normal numbers | Khi mô phỏng nhiễu, biến liên tục chuẩn | n: số phần tử |
| np.random.choice(a, size, replace) | Lấy mẫu từ mảng a | Sample from array | Khi chọn ngẫu nhiên mẫu | replace=True/False |
| np.random.seed(k) | Cố định kết quả ngẫu nhiên | Fix random state | Khi cần tái lập kết quả mô phỏng | k: số nguyên seed |

**🧮 3. Phân phối chuẩn (Normal Distribution)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| norm.pdf(x, loc, scale) | Hàm mật độ xác suất | Probability Density Function | Khi vẽ đường cong phân phối chuẩn | loc: trung bình μ, scale: độ lệch chuẩn σ |
| norm.cdf(x, loc, scale) | Hàm phân phối tích lũy | Cumulative Distribution Function | Khi tính P(X ≤ x) | — |
| norm.ppf(p, loc, scale) | Phân vị ngược | Percent Point Function (inverse CDF) | Khi tìm giá trị tại xác suất p | — |
| norm.rvs(loc, scale, size) | Sinh ngẫu nhiên từ phân phối chuẩn | Random Variates | Khi mô phỏng dữ liệu chuẩn | size: số phần tử |
| norm.fit(data) | Ước lượng tham số μ,σ từ dữ liệu | Estimate mean & std | Khi dữ liệu chưa biết tham số | — |

**Công thức PDF:**

**🧮 4. Phân phối đều (Uniform Distribution)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| uniform.pdf(x, loc, scale) | Hàm mật độ đều | Uniform density | Khi dữ liệu phân bố đều trên [a,b] | loc=a, scale=b−a |
| uniform.cdf(x, loc, scale) | Xác suất tích lũy | Cumulative function | Khi tính P(X≤x) | — |
| uniform.rvs(loc, scale, size) | Sinh dữ liệu ngẫu nhiên | Generate random uniform numbers | Khi mô phỏng ngẫu nhiên đều | — |

**Công thức PDF:**

**🧮 5. Phân phối nhị thức (Binomial Distribution)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| binom.pmf(k, n, p) | Xác suất rời rạc tại k | Probability Mass Function | Khi tính P(X=k) trong n phép thử | n: số lần, p: xác suất thành công |
| binom.cdf(k, n, p) | Xác suất P(X≤k) | Cumulative distribution | Khi cần xác suất tích lũy | — |
| binom.rvs(n, p, size) | Sinh ngẫu nhiên số lần thành công | Random samples | Khi mô phỏng thí nghiệm nhị thức | — |

**Công thức:**

**🧮 6. Phân phối Poisson (Poisson Distribution)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| poisson.pmf(k, mu) | Xác suất tại k | Probability of k occurrences | Khi mô phỏng số sự kiện trong thời gian xác định | mu: λ – trung bình |
| poisson.rvs(mu, size) | Sinh dữ liệu ngẫu nhiên | Generate random counts | Khi mô phỏng số lượt, số lỗi,... | — |
| poisson.cdf(k, mu) | Xác suất tích lũy | Cumulative probability | Khi cần P(X≤k) | — |

**Công thức:**

**🧮 7. Phân phối mũ (Exponential Distribution)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| expon.pdf(x, scale) | Hàm mật độ | Probability density | Khi mô phỏng thời gian giữa các sự kiện | scale = 1/λ |
| expon.cdf(x, scale) | Xác suất tích lũy | Cumulative probability | Khi tính P(X≤x) | — |
| expon.rvs(scale, size) | Sinh dữ liệu ngẫu nhiên | Random samples | Khi mô phỏng thời gian chờ | — |

**Công thức:**

**🧮 8. Các hàm mô tả và kiểm tra dữ liệu phân phối**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| .mean() | Trung bình | Mean | Xem giá trị kỳ vọng | — |
| .std() | Độ lệch chuẩn | Standard deviation | Mức độ phân tán | — |
| .var() | Phương sai | Variance | Mức độ biến thiên | — |
| .median() | Trung vị | Median | Khi dữ liệu lệch | — |
| plt.hist(data, bins, density=True) | Histogram | Plot data distribution | So sánh với PDF | — |
| plt.plot(x, pdf) | Đường cong lý thuyết | Plot PDF curve | Để đối chiếu thực nghiệm và lý thuyết | — |

**🎲 9. Mô phỏng & ước lượng (Simulation & Estimation)**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Tham số** |
| --- | --- | --- | --- | --- |
| np.mean(sample) | Tính trung bình mẫu | Sample mean | Ước lượng trung bình quần thể | — |
| np.std(sample, ddof=1) | Độ lệch chuẩn mẫu | Sample std | Ước lượng σ từ mẫu | ddof=1 cho mẫu |
| (sample >= value).sum()/len(sample) | Tỷ lệ phần tử thỏa điều kiện | Proportion of condition | Khi tính xác suất thực nghiệm | value: ngưỡng |
| plt.hist(samples, bins=30, density=True) | Histogram xác suất | Probability histogram | So sánh phân phối thực nghiệm | — |

**🔍 10. Hàm đặc biệt trong SciPy Stats**

| **Lệnh** | **Giải thích (Việt)** | **English** | **Khi nào dùng** | **Ghi chú** |
| --- | --- | --- | --- | --- |
| stats.describe(data) | Tóm tắt thống kê nhanh | Quick statistical summary | Khi cần toàn bộ đặc trưng dữ liệu | Trả về count, mean, variance, min, max |
| stats.probplot(data, plot=plt) | Biểu đồ Q–Q | Quantile-Quantile plot | Khi kiểm tra phân phối chuẩn | plot=plt để vẽ |
| stats.ttest\_1samp(data, popmean) | Kiểm định trung bình một mẫu | One-sample t-test | Khi kiểm tra giả thuyết về trung bình | popmean: giá trị so sánh |
| stats.ttest\_ind(a,b) | Kiểm định 2 mẫu độc lập | Independent t-test | So sánh 2 nhóm độc lập | — |
| stats.shapiro(data) | Kiểm định chuẩn | Shapiro-Wilk test | Khi kiểm tra dữ liệu có chuẩn không | — |

**📊 11. Các khái niệm quan trọng (Công thức & Giải thích)**

| **Ký hiệu** | **Công thức** | **Ý nghĩa** | **English Description** |
| --- | --- | --- | --- |
| Kỳ vọng |  | Trung bình | Expected value |
| Phương sai |  | Mức độ biến thiên | Variance |
| Độ lệch chuẩn σ |  | Độ phân tán | Standard deviation |
| Sai số chuẩn SE |  | Độ dao động của trung bình | Standard error |
| Luật số lớn | khi | Trung bình mẫu tiệm cận trung bình thật | Law of Large Numbers |
| Định lý giới hạn trung tâm | → chuẩn khi n lớn | Phân phối trung bình mẫu gần chuẩn | Central Limit Theorem |