

Histogram plot

Đây là loại biểu đồ phân bố tần suất, có dạng biểu đồ hình cột, mỗi cột đại diện cho phạm vi giá trị thay vì một giá trị duy nhất. Các bước để vẽ biểu đồ histogram:

- Khởi tạo bin dựa vào phạm vi của dữ liệu cần biểu diễn
- Chia data vào các bin
- Đếm số giá trị thỏa mãn thuộc mỗi bin

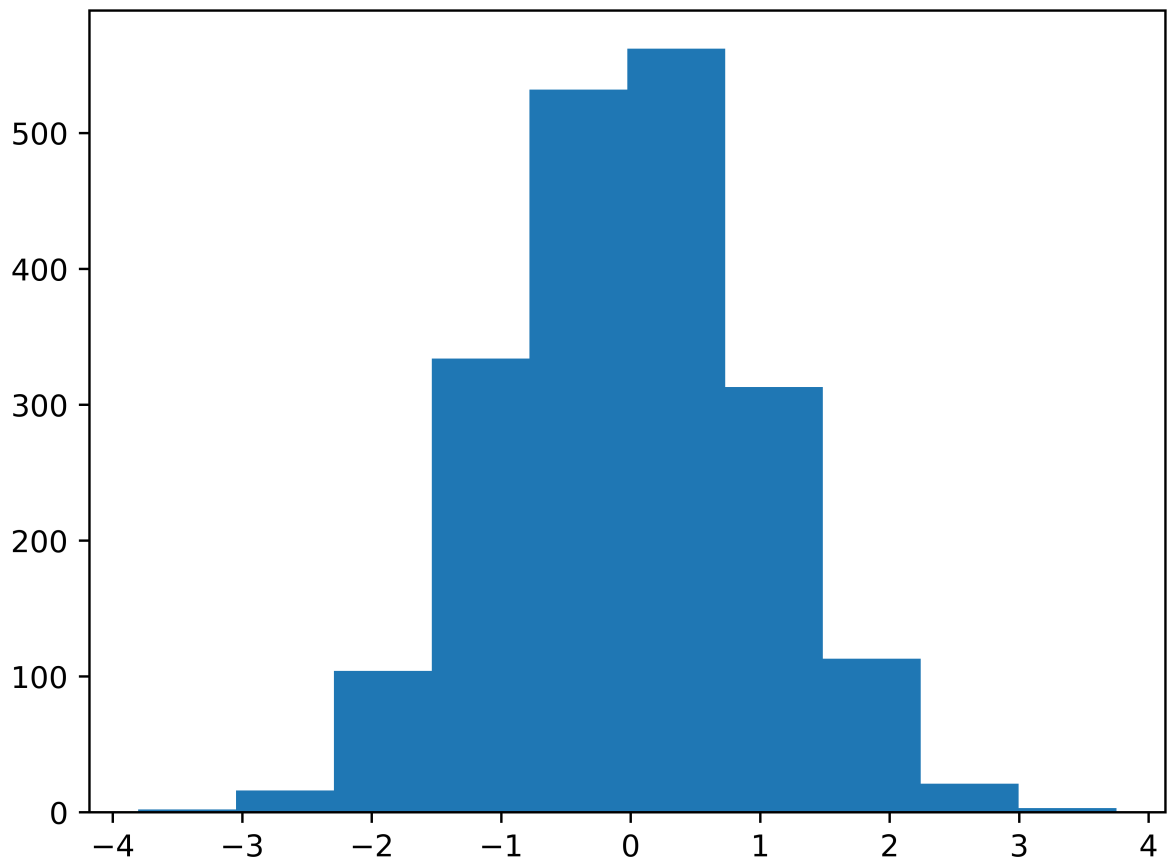
Hàm hist() trong matplotlib sử dụng để vẽ biểu đồ histogram, trong hàm này có các thuộc tính:

- x: dữ liệu đầu vào dạng list hoặc mảng numpy
- bins: có 2 loại dữ liệu truyền vào được chấp nhận
 - bins là list số nguyên sẽ định nghĩa độ rộng của mỗi bin
 - bins là list các list sẽ định nghĩa giá trị đầu và cuối các khoảng tương ứng của bin
- histtype: style của histogram gồm {'bar', 'barstacked', 'step', 'stepfilled'}, mặc định là bar
- density: giá trị true/false, mặc định là false. Lựa chọn có normalized theo một hàm mật độ xác suất
- cumulative: giá trị string tương ứng 'True'/'-1', nếu True, histogram vẽ sẽ cộng dồn các bin có kích thước nhỏ hơn

```
In [9]: import matplotlib.pyplot as plt
import numpy as np

rng = np.random.RandomState(42)
data = np.random.randn(2000)
fig, axes = plt.subplots(dpi=800)
# axes.hist(data, cumulative='True')
axes.hist(data)
```

```
Out[9]: (array([ 2., 16., 104., 334., 532., 562., 313., 113., 21., 3.]),
array([-3.80226987, -3.04690034, -2.29153081, -1.53616128, -0.78079176,
        -0.02542223,  0.7299473 ,  1.48531682,  2.24068635,  2.99605588,
         3.75142541]),
<BarContainer object of 10 artists>)
```



Vẽ nhiều histogram trên cùng một figure sử dụng `subplots()`, tùy chỉnh thêm kích thước và màu sắc

```
In [8]: import matplotlib.pyplot as plt
import numpy as np

data = np.random.randn(2000)

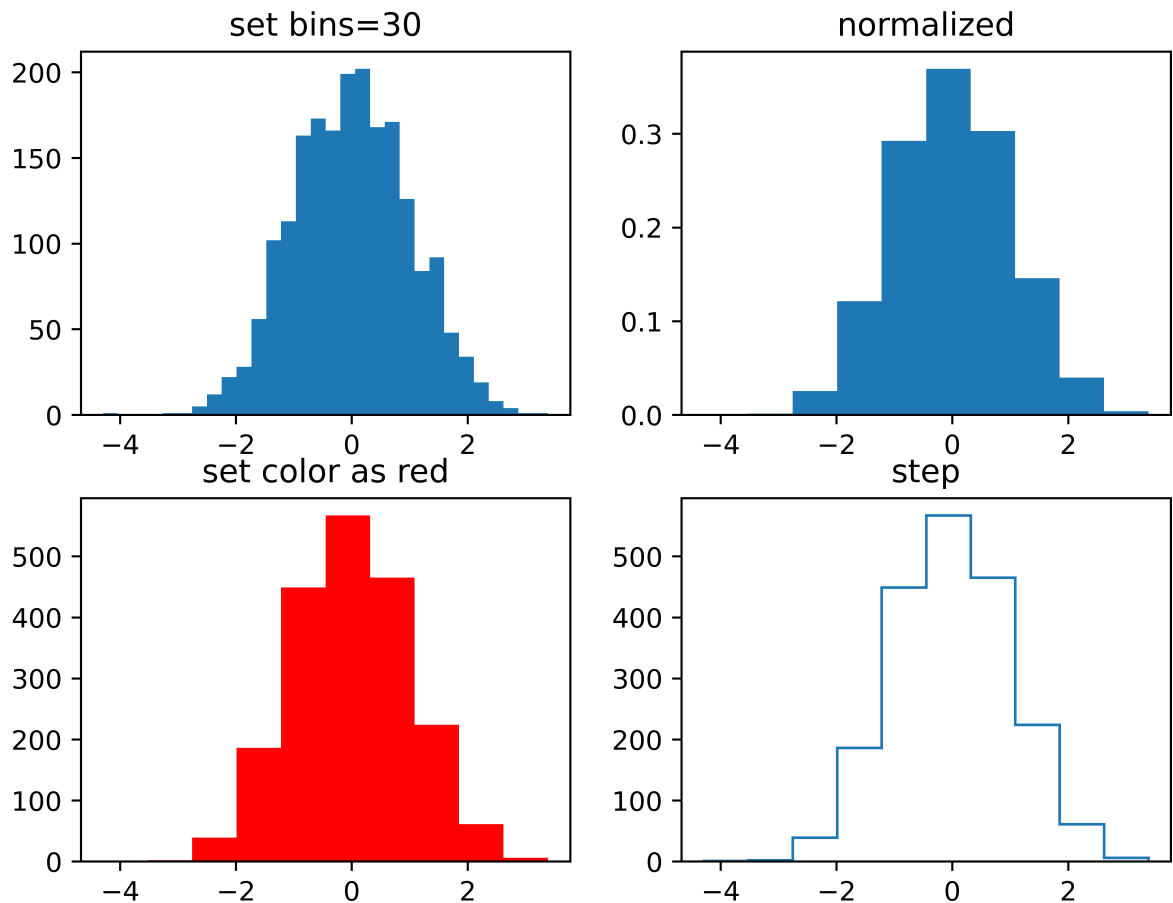
fig, axes = plt.subplots(nrows=2, ncols=2, dpi=800)
plt.tight_layout()
axes[0][0].hist(data, bins=30)
axes[0][0].set_title("set bins=30")

axes[0][1].hist(data, density=True)
axes[0][1].set_title("normalized")

axes[1][0].hist(data, color="r")
axes[1][0].set_title("set color as red")

axes[1][1].hist(data, histtype='step')
axes[1][1].set_title("step")
```

```
Out[8]: Text(0.5, 1.0, 'step')
```



Một số ví dụ về histogram trong thực tế như biểu đồ thống kê thời tiết, thống kê phổ điểm thi đại học. Trong các ví dụ này, thường có thêm đường cong uốn theo các cột biểu đồ để biểu diễn chất lượng của dữ liệu

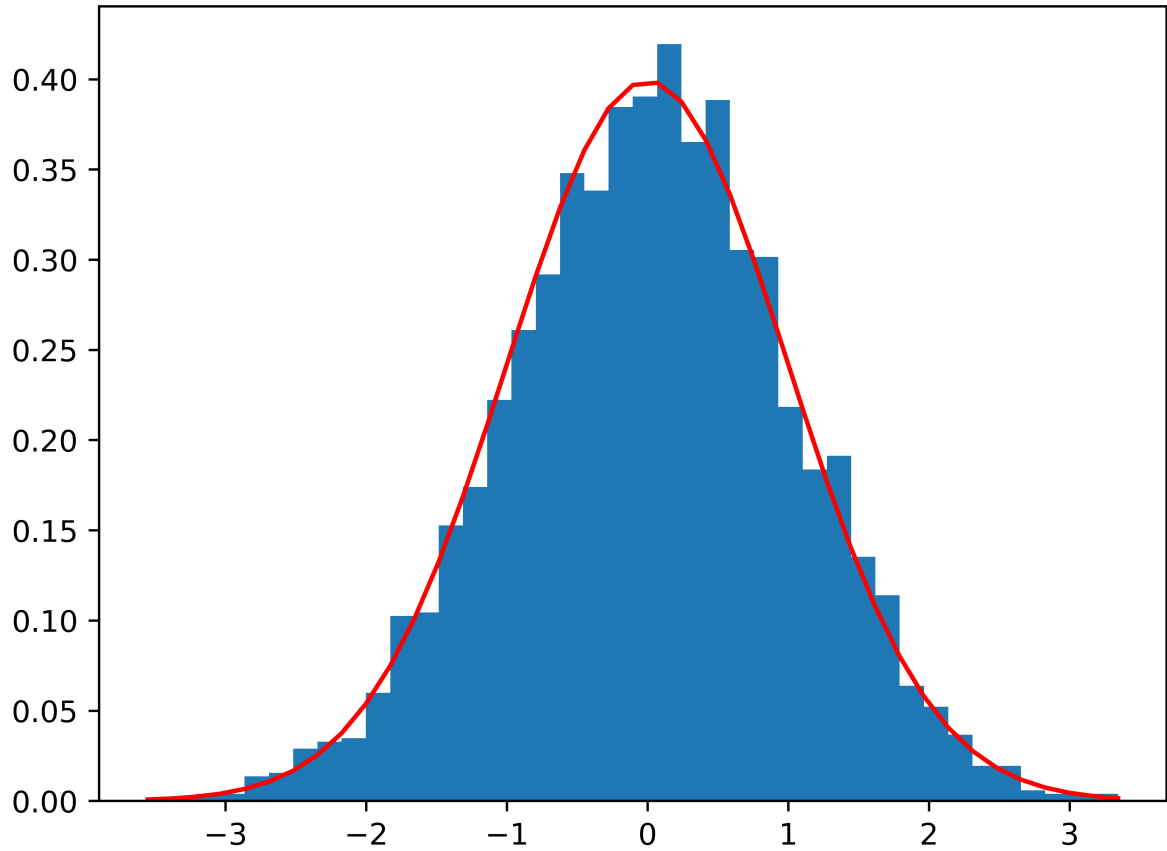
```
In [13]: import matplotlib.pyplot as plt
import numpy as np

sigma = 1
mu = 0
fig, axes = plt.subplots(dpi=800)
data = np.random.normal(mu, sigma, 3000) # khởi tạo mảng các giá trị ngẫu nhiên

n, bins, _ = axes.hist(data, bins=40, density=True)

y = ((1 / (np.sqrt(2 * np.pi) * sigma)) * np.exp(-0.5 * (1 / sigma * (bins -
axes.plot(bins, y, '-', color='r'))
```

```
Out[13]: [<matplotlib.lines.Line2D at 0x7fe52a7bf8b0>]
```



In []: