Administrative Items

- Special office hours
  - With Jason
  - Tomorrow @ 11am (Zoom)
  - See D2L Events for Zoom coordinates
  - TA office hours Monday (?) / Tuesday next week

- Revision to HW1 (See yesterday's note on Piazza)

- Problem 1(e) guidance…

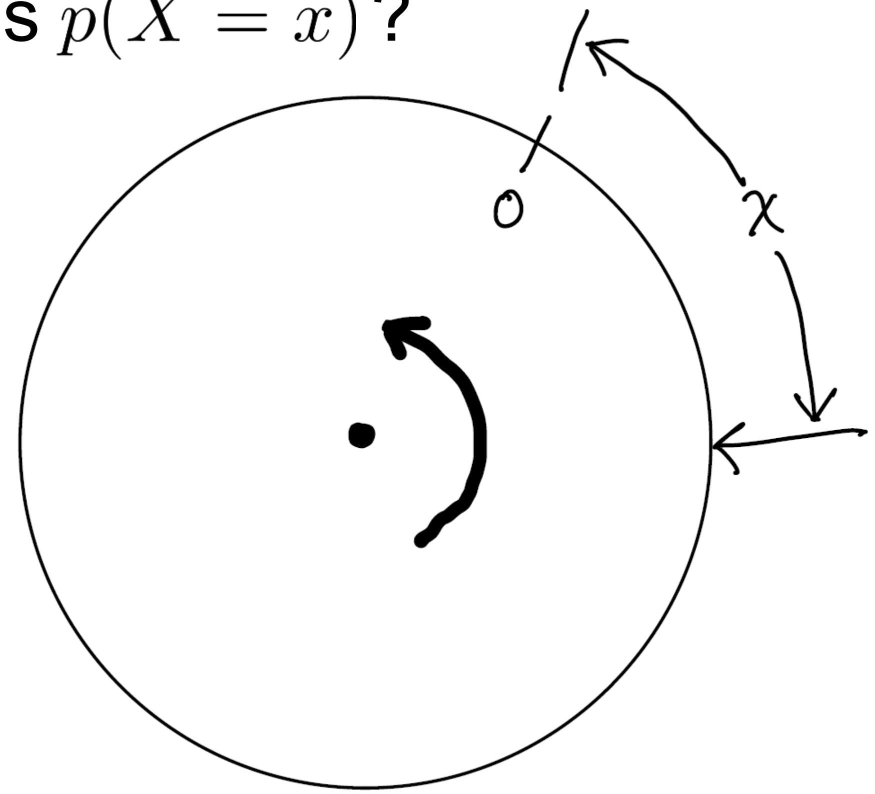
Outline

- Useful Discrete Distributions (+numpy.random)
- Continuous Probability
- Useful Continuous Distributions

Continuous Probability

Experiment Spin continuous wheel and measure X displacement from 0

Question Assuming uniform probability, what is $p(X = x)$?



First, recall axioms of probability…

1. For any event E, $0 \leq P(E) \leq 1$
2. $P(\Omega) = 1 \text{ and } P(\emptyset) = 0$
3. For any *finite* or *countably infinite* sequence of <u>pairwise mutually disjoint events</u> $E_1, E_2, E_3, \ldots$

$$P\Big(\bigcup_{i \geq 1} E_i\Big) = \sum_{i \geq 1} P(E_i)$$

Sample space Ω is all points (real numbers) in $[0, 1)$

Continuous Probability

- Let $p(X = x) = \pi$ be the probability of any single outcome
- Let $S(k)$ be set of any k *distinct* points in $[0, 1)$ then,

$$P(x \in S(k)) = k\pi$$

- Since $0 < P(x \in S(k)) < 1$ by axioms of probability, $k\pi < 1$ for any k
- Therefore: $\pi = 0$ and $P(x \in S(k)) = p(X = x) = 0$

What does this mean?

- Let E be event that $x \in S(k)$
- In infinite sample space, an event may be **possible** but have zero "probability"
- Since $P(\bar{E}) = 1 - P(E) = 1$ events may have "probability" 1 but **not certain**

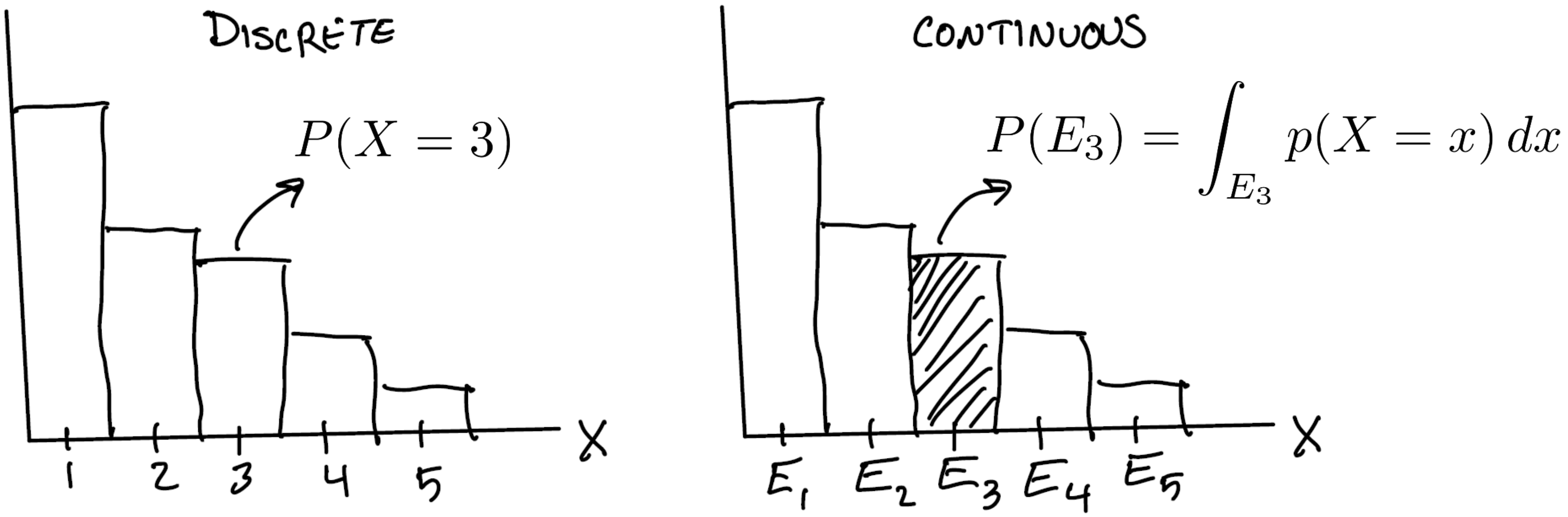
Assign probability to intervals, not individual values

Continuous Probability

We could just accept this oddity…

…or we could try to convince ourselves that it is sensible.

Continuous Probability



- Events represented as intervals $a \leq X < b$ with probability,

$$P(a \leq X < b) = \int_a^b p(X = x)\,dx$$

- Specific outcomes have zero probability $P(X = 0) = P(x \leq X < x) = 0$
- But may have nonzero *probability density* $p(X = x)$

Continuous Probability

Definition *The cumulative distribution function (CDF) of a real-valued continuous RV X is the function given by,*

$$F(x) = P(X \leq x)$$

Different ways to represent probability of interval, CDF is just a convention.

- Can easily measure probability of closed intervals,

$$P(a \leq X < b) = F(b) - F(a)$$

- If *X* is differentiable then,

Fundamental Theorem of Calculus

$$f(x) = \frac{dF(x)}{dx} \quad \text{and} \quad F(t) = \int_{-\infty}^{t} f(x)\, dx$$

Where $f(x)$ is the *probability density function* (PDF)

- Typically use shorthand P for CDF and p for PDF instead of F and f

Continuous Probability

Most definitions for discrete RVs hold, replacing sum with integral…

Law of Total Probability for continuous distributions,

$$p(x) = \int_{\mathcal{Y}} p(x, y)\, dy$$

…and replacing PMF with PDF…

Conditional PDF:

$$p(X \mid Y) = \frac{p(X,Y)}{p(Y)} = \frac{p(X,Y)}{\int p(x,Y)\, dx}$$

Probability Chain Rule:

$$p(X, Y) = p(Y)p(X \mid Y)$$

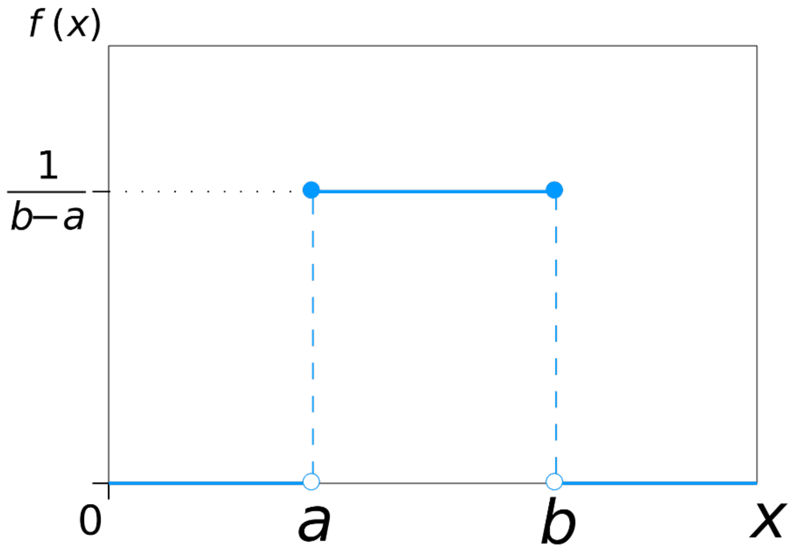
Outline

- Useful Discrete Distributions (+numpy.random)
- Continuous Probability
- Useful Continuous Distributions

Useful Continuous Distributions

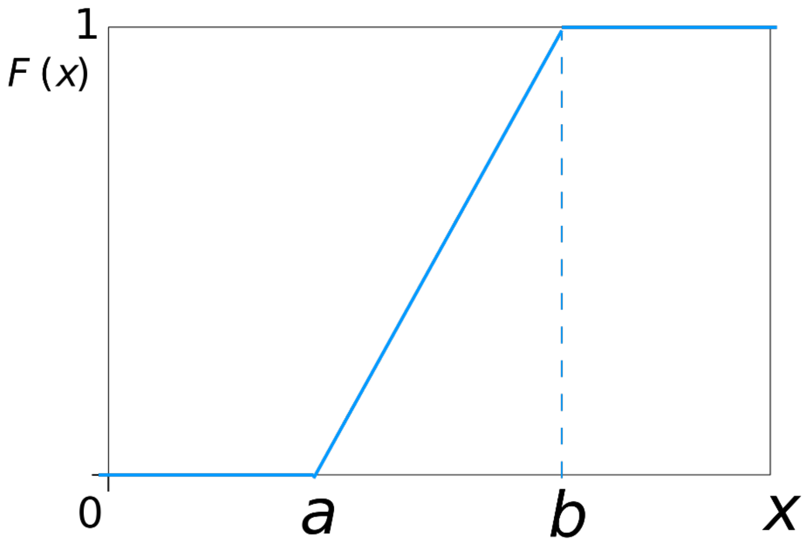
Uniform distribution on interval $[a, b]$,

$$p(x) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{if } b \leq x \end{cases} \qquad P(X \leq x) = \begin{cases} 0 & \text{if } x \leq a, \\ \frac{x-a}{b-a} & \text{if } a \leq x \leq b, \\ 1 & \text{if } b \leq x \end{cases}$$



Say that $X \sim U(a, b)$ the average is,

$$\mathbf{E}[X] = \frac{b+a}{2}$$



Suppose $X \sim U(0, 1)$ and we are told $X \leq \frac{1}{2}$ what is the conditional distribution?

$$P(X \leq x \mid X \leq \tfrac{1}{2}) = U(0, \tfrac{1}{2})$$

Holds generally: Uniform closed under conditioning

numpy.random.uniform

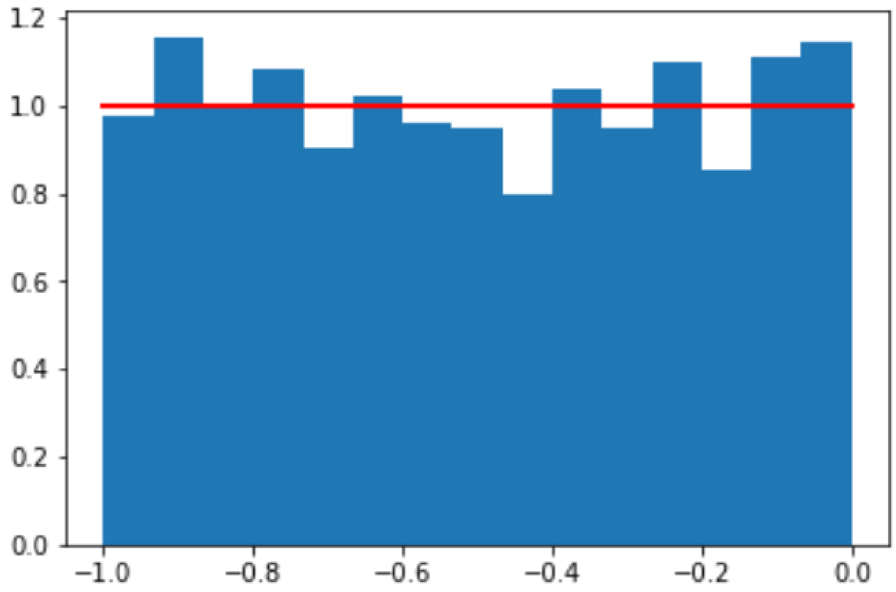
numpy.random.uniform(*low=0.0*, *high=1.0*, *size=None*)

Draw samples from a uniform distribution.

Samples are uniformly distributed over the half-open interval `[low, high)` (includes low, but excludes high). In other words, any value within the given interval is equally likely to be drawn by **uniform**.

Example Drawn 1,000 samples from a uniform on [-1,0),

```
a = -1
b = 0
N = 1000
X = np.random.uniform(a,b,N)
count, bins, ignored = plt.hist(X, 15, density=True)
plt.plot(bins, np.ones_like(bins), linewidth=2, color='r')
plt.show()
```

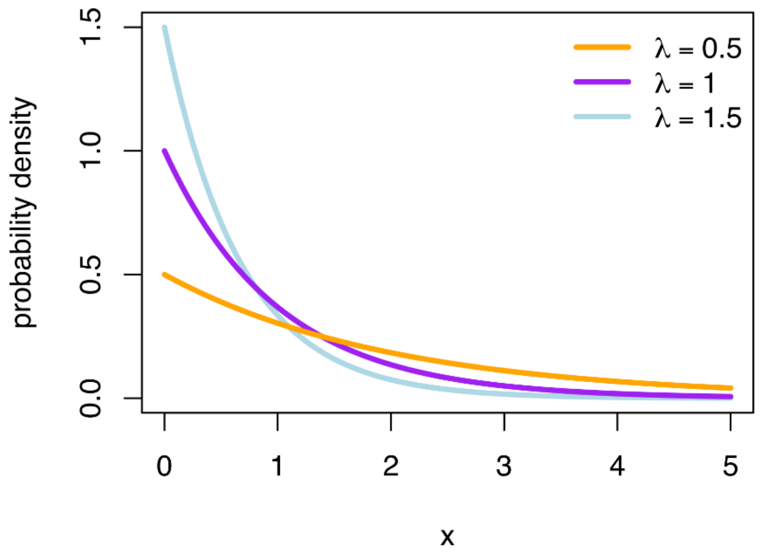


Useful Continuous Distributions

Exponential distribution with scale λ,

$$p(x) = \lambda e^{-\lambda x} \qquad\qquad P(x) = 1 - e^{-\lambda x}$$

for X>0. Expected value given by, $\mathbf{E}[X] = \frac{1}{\lambda}$



Useful properties

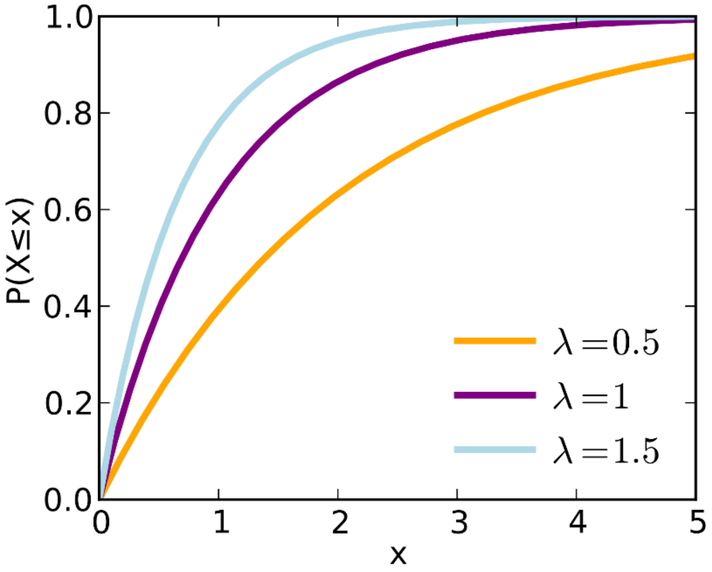
- **Closed under conditioning** If $X \sim \text{Exponential}(\lambda)$ then,

$$P(X \geq s + t \mid X \geq s) = P(X \geq t) = e^{-\lambda t}$$

- **Minimum** Let $X_1, X_2, \ldots, X_N$ be i.i.d. exponentially distributed with scale parameters $\lambda_1, \lambda_2, \ldots, \lambda_N$ then,

$$P(\min(X_1, X_2, \ldots, X_N)) = \text{Exponential}(\textstyle\sum_i \lambda_i)$$

- Time between $\text{Poisson}(\lambda)$ arrivals is distributed as $\text{Exponential}(\frac{1}{\lambda})$



numpy.random.exponential

numpy.random.exponential(*scale=1.0, size=None*)

Draw samples from an exponential distribution.

Its probability density function is

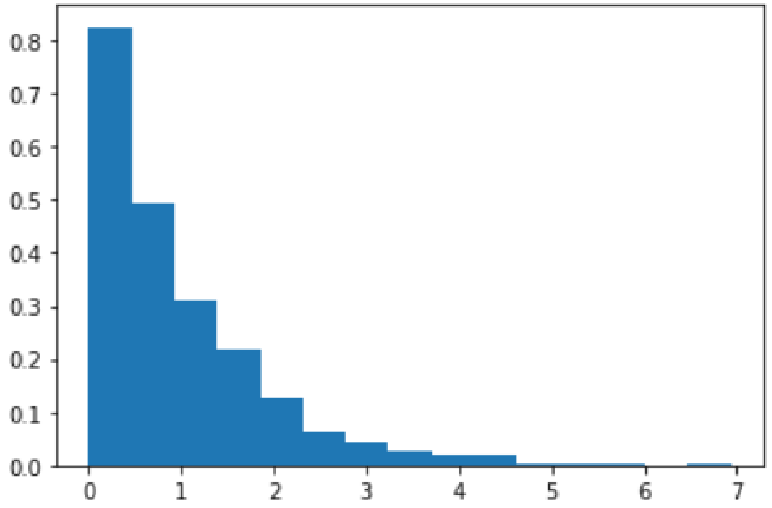
$$f(x; \frac{1}{\beta}) = \frac{1}{\beta} \exp(-\frac{x}{\beta}),$$

for `x > 0` and 0 elsewhere. β is the scale parameter, which is the inverse of the rate parameter $\lambda = 1/\beta$. The rate parameter is an alternative, widely used parameterization of the exponential distribution [3].

The exponential distribution is a continuous analogue of the geometric distribution. It describes many common situations, such as the size of raindrops measured over many rainstorms [1], or the time between page requests to Wikipedia [2].

Example Draw 1,000 samples from exponential with $\lambda = 1.0$

```
lam = 1.0
N = 1000
X = np.random.exponential(lam, N)
count, bins, ignored = plt.hist(X, 15, density=True)
plt.show()
```

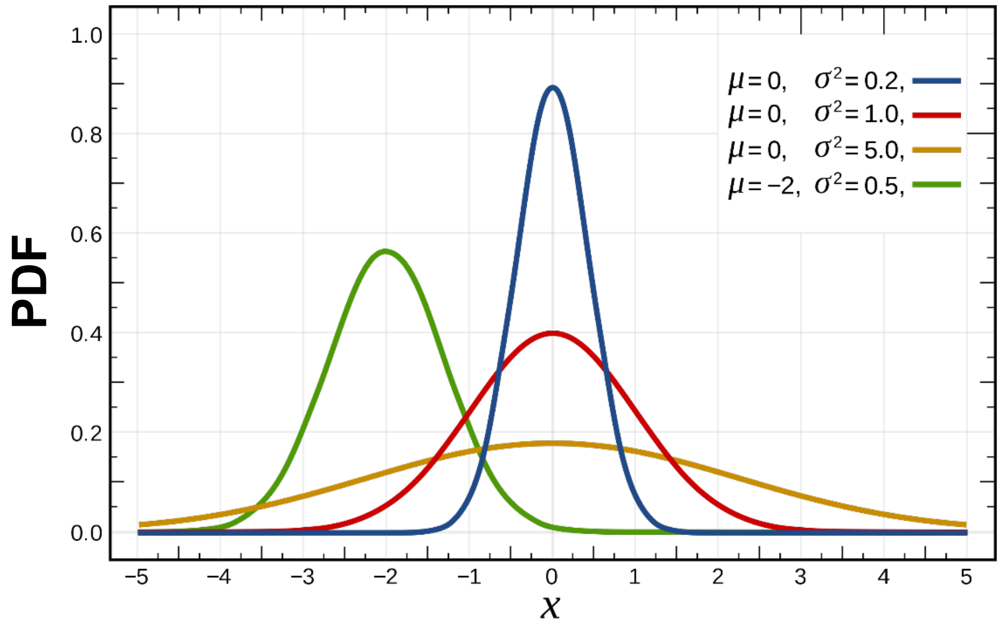


Useful Continuous Distributions

Gaussian (a.k.a. Normal) distribution with mean mean (location) μ and variance (scale) σ^2 parameters,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2}(x-\mu)^2/\sigma^2$$

We say $X \sim \mathcal{N}(\mu, \sigma^2)$.



Useful Properties

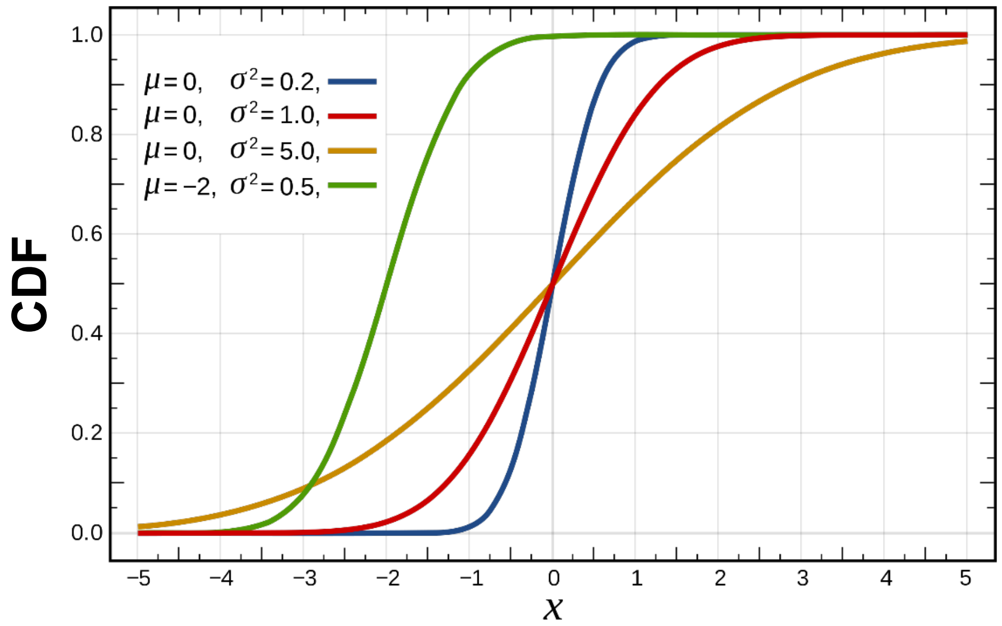
- Closed under additivity:

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

- Closed under linear functions (a and b constant):

$$aX + b \sim \mathcal{N}(a\mu_x + b, a^2\sigma_x^2)$$



numpy.random

numpy.random.normal

numpy.random.normal(*loc=0.0, scale=1.0, size=None***)**

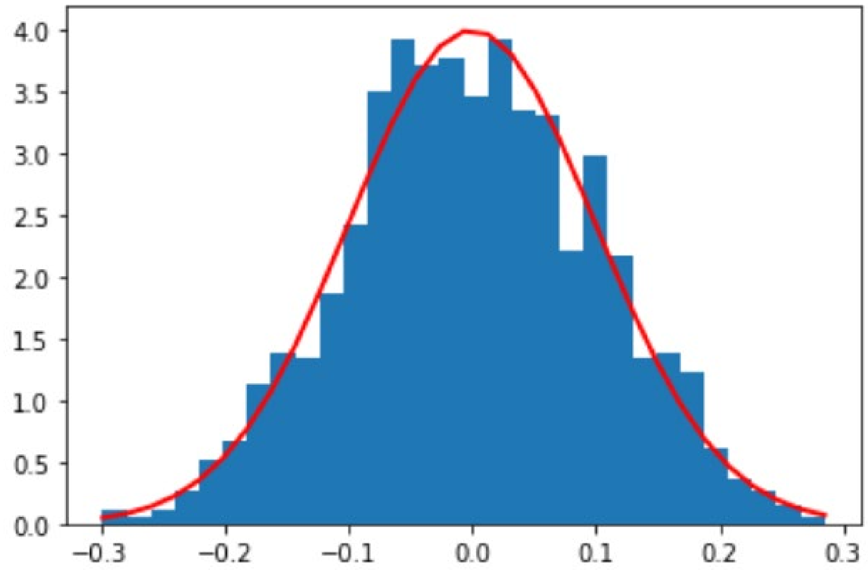
Draw random samples from a normal (Gaussian) distribution.

The probability density function of the normal distribution, first derived by De Moivre and 200 years later by both Gauss and Laplace independently [2], is often called the bell curve because of its characteristic shape (see the example below).

The normal distributions occurs often in nature. For example, it describes the commonly occurring distribution of samples influenced by a large number of tiny, random disturbances, each with its own unique distribution [2].

Example Sample zero-mean gaussian with STDEV 0.1,

```
mu, sigma = 0, 0.1 # mean and standard deviation
X = np.random.normal(mu, sigma, 1000)
count, bins, ignored = plt.hist(X, 30, density=True)
plt.plot(bins, 1/(sigma * np.sqrt(2 * np.pi)) *
         np.exp( - (bins - mu)**2 / (2 * sigma**2) ),
         linewidth=2, color='r')
plt.show()
```



numpy.random

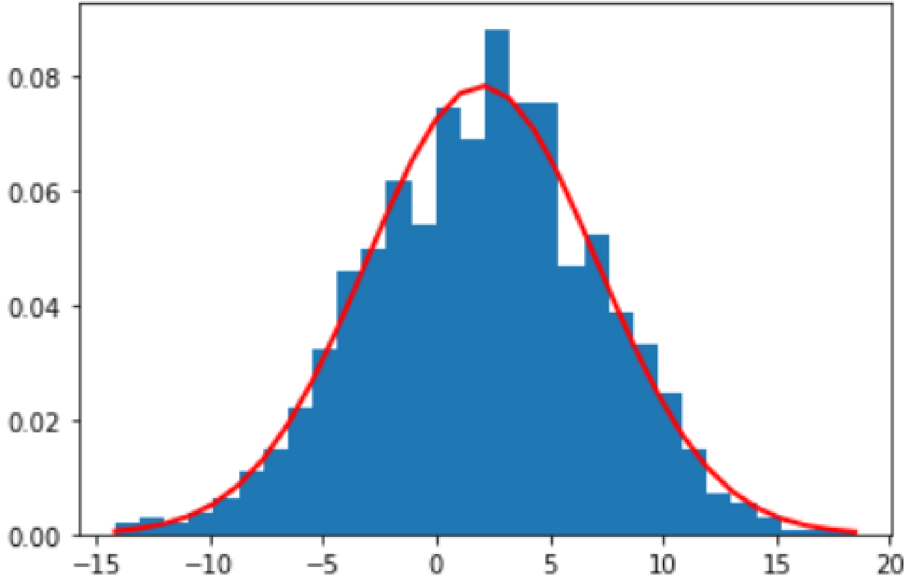
Gaussians are closed under additivity

Example *Add two Gaussian RVs,*

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

```
mu_x, sigma_x = 0, 1
mu_y, sigma_y = 2, 5
X = np.random.normal(mu_x, sigma_x, 1000)
Y = np.random.normal(mu_y, sigma_y, 1000)
Z = X+Y
count, bins, ignored = plt.hist(Z, 30, density=True)
mu_z = mu_x + mu_y
sig_z_sq = sigma_x**2 + sigma_y**2
plt.plot(bins, 1/(np.sqrt(sig_z_sq * 2 * np.pi)) *
         np.exp( - (bins - mu_z)**2 / (2 * sig_z_sq) ),
         linewidth=2, color='r')
plt.show()
```



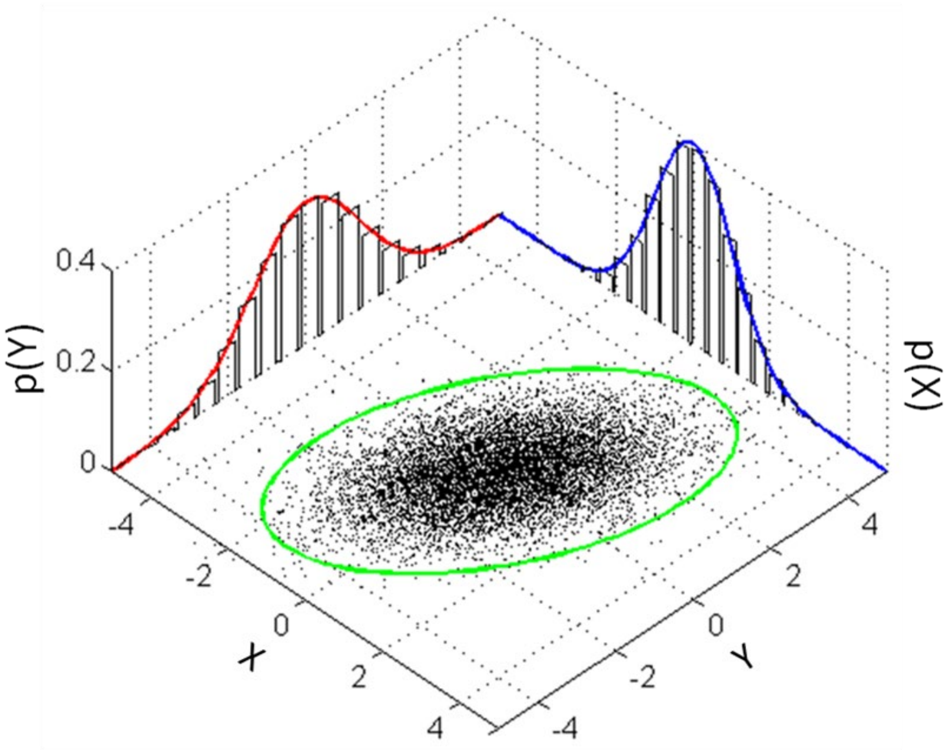
Property extends to a sequence of Gaussian RVs,

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \qquad \sum_i X_i \sim \mathcal{N}(\cdot)$$

Useful Continuous Distributions

Multivariate Gaussian On RV $X \in \mathcal{R}^d$ with mean $\mu \in \mathcal{R}^d$ and positive semidefinite covariance matrix $\Sigma \in \mathcal{R}^{d \times d}$,

$$p(x) = |2\pi\Sigma|^{-1/2} \exp -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)$$



Moments given by parameters directly.

Useful Properties

- Closed under additivity (same as univariate case)
- Closed under linear functions,

$$AX + b \sim \mathcal{N}(A\mu_x + b, A\Sigma A^T)$$

Where $A \in \mathcal{R}^{m \times d}$ and $b \in \mathcal{R}^m$ (output dimensions may change)

- Closed under conditioning and marginalization

Will discuss Gaussians a lot more when we cover exponential families

numpy.random

numpy.random.multivariate_normal

numpy.random.multivariate_normal(*mean*, *cov*[, *size*, *check_valid*, *tol*]**)**

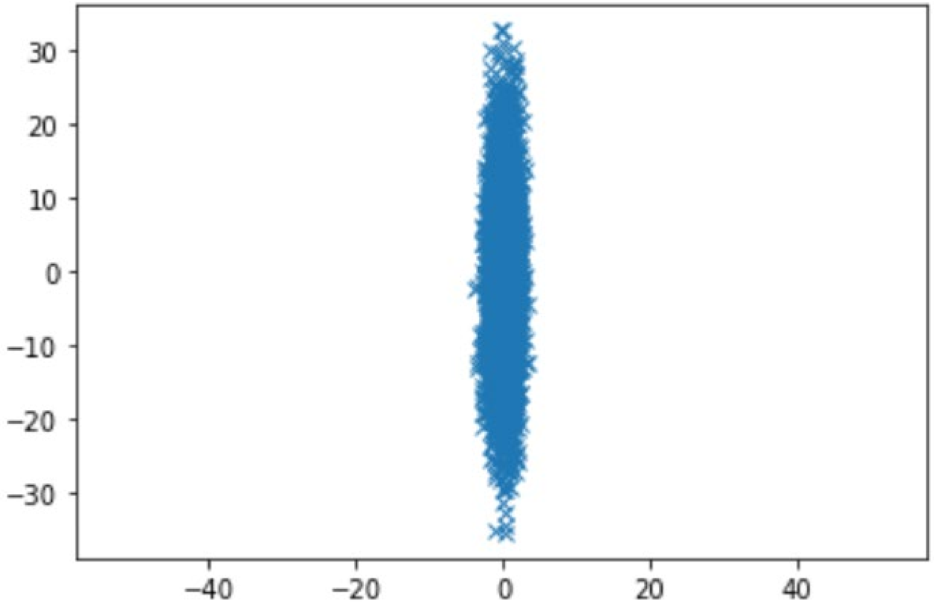
Draw random samples from a multivariate normal distribution.

The multivariate normal, multinormal or Gaussian distribution is a generalization of the one-dimensional normal distribution to higher dimensions. Such a distribution is specified by its mean and covariance matrix. These parameters are analogous to the mean (average or "center") and variance (standard deviation, or "width," squared) of the one-dimensional normal distribution.

Example Sample from zero-mean 2D (bivariate) Gaussian with covariance

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}$$

```
mean = [0, 0]
cov = [[1, 0], [0, 100]]  # diagonal covariance
x, y = np.random.multivariate_normal(mean, cov, 5000).T
plt.plot(x, y, 'x')
plt.axis('equal')
plt.show()
```

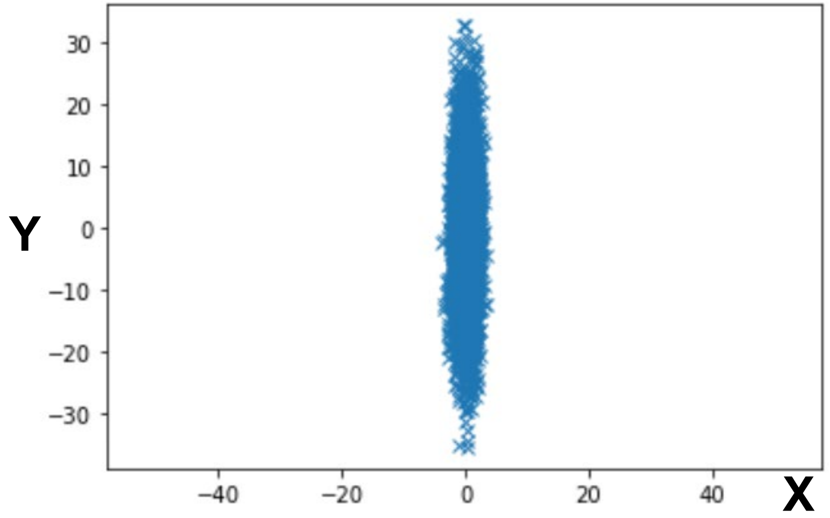


numpy.random

Multivariate Gaussians closed under marginalization

Example Bivariate Gaussian,

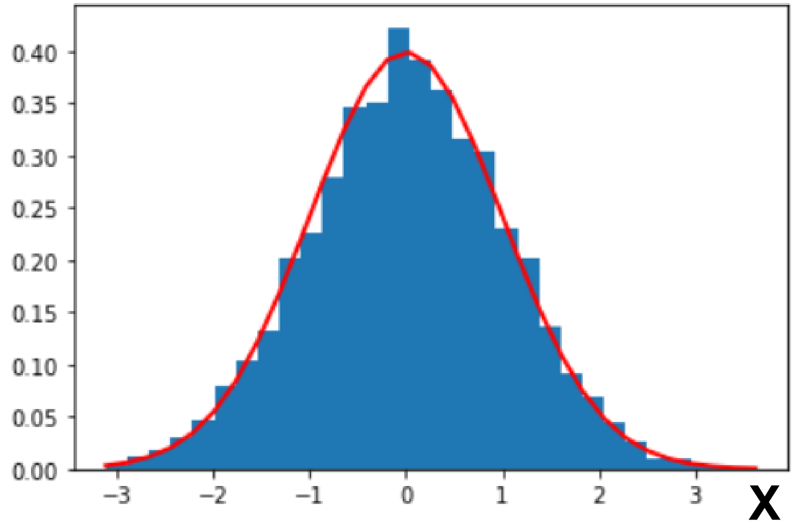
$$p(X,Y) = \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}\right)$$



Marginal moments are elements from joint,

$$p(X) = \int p(X,y)\,dy = \mathcal{N}(\mu_x, \sigma_x^2)$$

```
sig_x_sq = 1
mu_x = 0
count, bins, ignored = plt.hist(x, 30, density=True)
plt.plot(bins, 1/(np.sqrt(sig_x_sq * 2 * np.pi)) *
         np.exp( - (bins - mu_x)**2 / (2 * sig_x_sq) ),
         linewidth=2, color='r')
plt.show()
```



Recap

Useful discrete distributions

- Bernoulli → "Coinflip Distribution"
- Binomial → Multiple Bernoulli draws
- Geometric → How long do I have to wait to see "heads"?
- Categorical / Multinomial → One / Many die rolls
- Poisson → "Arrivals" in a fixed time

Continuous probability

- P(X=x) = 0 <u>does not mean impossible</u>
- 1 – P(X=x) = 1 <u>does not mean certain</u>
- Probabilities assigned to *intervals* via CDF P(X > x)
- PDF measures probability *density* of single points p(X=x) >= 0

Useful continuous distributions

- Exponential → Time between Poisson arrivals
- Univariate / Multivariate Gaussian → Probably most ubiquitous distribution
- There are a lot more we will touch on later in the course…