Week 2

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm # imports normal distribution
```

First Session

Exercise 1 Probability for Continuous Variables

For your hobby deep-learning projects, you want to buy an expensive graphics processing unit (GPU) to accelerate your computations.

As an aware consumer, you carefully investigate the risk of the GPU becoming defective and the procedure for having it replaced for free, before deciding on the purchase.

You learn that in Denmark, private consumers are protected by a legal guarantee when buying goods, that covers free repairs or replacements. The duration of the legal guarantee is two years. In case of a legal twist between the seller and consumer, the law states that: For the first six month after the delivery, the seller must prove that the item was not defective. For the rest of the guarantee period the consumer must prove that the product was defective.

You learn that the time x at which the product becomes defective is exponentially distributed with probability density function,

$$p(x) = \frac{1}{\beta} e^{-\frac{1}{\beta}x},$$

where x is a real number measuring *years after purchase* and the scale parameter β is known $\beta = 2$.

1. Plot $p(x)$ when $x \in [0, 10]$.
2. What is the probability that the product becomes defective within the two-year guarantee but after the initial six months?
3. If the product still works after two years, what is the (conditional) probability that it will work for at least two more years?
4. What is the probability that the product continues to work for at least the first ten years?

Hints

Notice that $p(x)$ is supported in the interval $0 \leq x < \infty$.

In python, the value e^a can be computed using the exponential function `np.exp(a)`.

Exercise 2 Normals and Standard Deviation

When computing integrals of probability densities, it's often convenient to work with the cumulative distribution function (CDF) of the random variable instead.

$$F(x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x \qquad \text{(cumulative distribution function)}$$

Because of the fundamental theorem of calculus, $\int_a^b p(x)\,\mathrm{d}x = F(b) - F(a)$, knowing the CDF allows us to easily calculate probabilities.

In the `scipy.stats` package you can find the `norm` class which can be used to define a normal distribution, e.g. `norm(0,1)` defines the standard normal distribution[1].

1. Figure out how to call the normal CDF using `norm`'s methods. For the three distributions $\mathcal{N}(0,1)$, $\mathcal{N}(-2,3)$, and $\mathcal{N}(0,10)$.
   a) The probability of sampling a point in $[-0.1, 0.1]$.
   b) The probability of being within three standard deviations of the mean.
   c) The probability of one hundred independent samples all landing within three standard deviations of the mean.

Exercise 3 Histograms and Densities

Often, the density of a random variable is not known beforehand. If we have collected samples of the random variable, we can try to approximate the density.

1. Recall that all probability density functions $p(x)$ integrate to 1.
   a) What is the integral of a function that is constant $f(x) = h$ over an interval[2] of width w, and 0 everywhere else?
   b) If $p(x) = Cf(x)$ is a uniform probability density function on the same interval, and $f(x)$ is the function from the previous question, what should the constant C be so that the two match?
   c) Now take a function consisting of K constant functions of height h_k on intervals B_k each of width w_k, that is,

$$f(x) = \sum_{k=1}^{K} \mathbb{1}[x \in B_k]\, h_k \qquad \text{(staircase function)}$$

[1] Pay attention to the documentation: Scipy does not parameterize the normal using mean and variance.
[2] e.g. $[0, w]$, although note that it does not matter

where $\mathbb{1}[x \in B_k]$ is 1 if x is in B_k and 0 otherwise. If $p(x) = Cf(x)$ is again a probability density function and $f(x)$ is the staircase function defined above, what should the constant C be?

d) If you draw a sample from the density $p(x)$ you just defined, what is the probability of the sample landing in each interval B_k?

A *histogram* is a type of plot that visualizes the distribution of a set of points $\{x_n\}_{n=1}^N$ by dividing them into non-overlapping intervals $B_k = [b_k, b_{k+1}]$ called *bins*, and then plotting each interval as a bar with height equal to the number of points that fell into the bin. Seen as a function, it is a staircase function like above with $w_k = |b_{k+1} - b_k|$ and $h_k = \sum_{n=1}^N \mathbb{1}[x_n \in B_k]$ (counting the number of elements that go into bin k).

2. For this question we want to investigate how a normalized histogram approximates the true probability density function. We will focus on the interval $[-5, 5]$ and assume that all the bins have the same width.

   Use the following code to define the bins.

```
nbins = # choose number of bins

#split the interval [-5,5] into bins
bin_borders = np.linspace(-5,5,nbins+1)
bin_width = 10./nbins
bin_centers = bin_borders[:-1]+0.5*bin_width
```

   a) Use `np.random.randn` to draw 10000 samples from a standard normal distribution $\mathcal{N}(0, 1)$.

   b) Compute the histogram of the samples for varying values of `nbins` using `np.histogram`. Normalize it so that it is a density. Try the values 10, 20, 100, and 1000 and plot each using `plt.bar` and compare with the true probability density function using the `norm` object. Comment on the differences in quality.

In the future, you can use `plt.hist(samples, bin_borders, density=True)` to plot a normalized histogram more conveniently.

Exercise 4 Expectations

Recall the following rules about expectations,

$$
\begin{aligned}
\mathbb{E}[aX + bY] &= a\,\mathbb{E}[X] + b\,\mathbb{E}[Y] && \text{(linearity)} \\
\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|Y]] && \text{(tower property)} \\
X, Y \text{independent} \Rightarrow \mathbb{E}[XY] &= \mathbb{E}[X]\,\mathbb{E}[Y] && \text{(independent expectations)} \\
Z = g(X) \Rightarrow \mathbb{E}[Z] &= \int p_X(x) g(x)\,\mathrm{d}x && \text{(LOTUS)}
\end{aligned}
$$

1. Convince yourself that

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbb{1}[X \in A]] \tag{1}$$

for any random variable X and some interval A.

2. Prove that

$$\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2. \tag{2}$$

3. Prove that if X and Y are independent,

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y), \tag{3}$$

where $\mathrm{Var}(Z) = \mathbb{E}\left[(Z - \mathbb{E}[Z])^2\right]$ is the variance of a random variable Z.

Exercise 5 Defective Device Test

In a factory, you have a device that tries to predict if a unit is defective by running a gauntlet of tests resulting in a final test score T. 10% of the units turn out to be defective. If the unit works as expected (call the event $W = 1$) it will score 1 on the test on average,

$$T|W = 1 \sim \mathcal{N}(1, 1). \tag{4}$$

If the unit is defective it will score -1 on average,

$$T|W = 0 \sim \mathcal{N}(-1, 1). \tag{5}$$

Our testing device assumes that things with negative scores are broken using the decision rule,

$$D(t) = \begin{cases} 1 & t > 0, \quad \text{(unit is working)} \\ 0 & t \leq 0, \quad \text{(unit is defective)} \end{cases} \tag{6}$$

1. We can imagine that it costs us 100 if we either discard a working unit or have to refund a broken unit that was sold, so for a random W and T the cost is
$$L(W,T) = 100\,\mathbb{1}[W \neq D(T)]\,. \tag{7}$$
Remember that $\mathbb{1}[\cdot]$ is the indicator function, which is 1 if the statement inside is true, and 0 otherwise. Convince yourself that L can also be written as,
$$L(W,T) = 100\,(\mathbb{1}[W = 0]\,\mathbb{1}[T > 0] + \mathbb{1}[W = 1]\,\mathbb{1}[T \leq 0])\,. \tag{8}$$

2. Use the tower property and the other rules to compute the expected cost
$$\mathbb{E}_{p(W,T)}[L(W,T)]\,. \tag{9}$$

Second Session

See the separate Jupyter notebook `exercises_week2_session2.ipynb`.