

Homework 3

Due October 3 at 11 pm

1. (Half life) The half life of a radioactive substance is a way to quantify how rapidly the substance decays. Given a fixed quantity of the substance, the half time is the time that it takes for it to be reduced to half (i.e. half of the radioactive particles have decayed). It is not immediately apparent why the time should be the same for any quantity. Here we'll show that it is (probabilistically) if the particles decay following an exponential distribution.

(a) Let \tilde{t} be a random variable with a pdf of the form

$$f_{\tilde{t}}(t) := \begin{cases} \lambda \exp(-\lambda t), & \text{if } t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where λ is a fixed constant. We define the half life $t_{1/2}$ as the number that satisfies $P(\tilde{t} > t_{1/2}) = 1/2$. Compute $t_{1/2}$ in terms of λ . Then explain intuitively why this is a reasonable definition for the half life.

We can calculate $t_{1/2}$ in terms of λ by plugging in $t_{1/2}$ into our random variable PDF, then integrate to get our cdf. First we integrate our pdf function:

$$P(\tilde{t} > t) = \int_{t_{1/2}}^{\infty} \text{PDF Function} = \int_{t_{1/2}}^{\infty} \lambda e^{-\lambda x} = e^{-\lambda t_{1/2}}$$

Then we can manipulate with algebra to get:

$$\frac{1}{2} = e^{-\lambda t_{1/2}} \rightarrow \frac{\ln(.5)}{-\lambda} = t_{1/2} \rightarrow \frac{\ln(2)}{\lambda} = t_{1/2}$$

This is a reasonable definition for half life and is intuitive makes as we calculated the probability that 50% of our matter would be gone. This makes sense our cdf function tells us the probability of an event up to a certain point and its calculated as $1 - e^{-\lambda(t-t_{1/2})}$. Since we set the other side to $\frac{1}{2}$ we effectively are asking the question, at what point does the probability equal 50% on each side? Which we have defined as $t_{1/2}$ and this works since we assumed all the particles have the same distribution.

- (b) Compute t such that $P(t_{1/2} < \tilde{t} < t) = 1/4$, and express it in terms of only $t_{1/2}$. Explain why the result is consistent with the intuitive meaning of half life.

$$P(t_{1/2} < \tilde{t} < t) = 1 - e^{-\lambda(t-t_{1/2})} = \frac{1}{2} - e^{-\lambda t} = \frac{1}{4} \rightarrow t_{1/4} = \frac{\ln(4)}{\lambda} \rightarrow t_{1/4} = 2 \times t_{1/2}$$

Again, this intuitively makes sense as we are asking the question, how long would it take for our matter to be halved, and then be halved again? The function is "memory-less" and therefore, after it is halved once, the time that it takes to get halved again does not change. So the time to have $t_{1/4}$ of the matter is twice the amount of time to half the matter once, or $2 \times t_{1/2}$

- (c) Compute $P(\tilde{t} > kt_{1/2})$ for any integer k . Again, explain why the result is consistent with the intuitive meaning of half life.

$$P(\tilde{t} > kt_{1/2}) \text{ where } k \in \mathbf{Z} = 1 - e^{-(\lambda kt_{1/2})} \rightarrow e^{\ln(2^k)} = \frac{1}{2^k}$$

The answer we derived is no surprise, as the definition of half-life is the time it takes for the mass of the radioactive quantity to be reduced by half, the time it takes to lose one half of its mass is independent of any amount of its mass it has lost until that point (it's "memory-less"). Therefore, our equation that represents half-life should be "memory-less" as well, and thus the exponential function we used to model the pdf is a great choice, as the pdf of the exponential function at time t given some time t_0 is equivalent to the pdf exponential function shifted over $(t - t_0)$ units .

2. (Measurements) You have access to the readings of a device that indicates whether a radioactive particle has decayed. However you do not get a continuous reading, you get a reading every second.
- (a) A reasonable model for the time the particle takes to decay is that it is a random variable with pdf

$$f_{\tilde{t}}(t) := \begin{cases} \lambda \exp(-\lambda t), & \text{if } t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where λ is a fixed constant. Taking into account that the measurement device rounds up the time and outputs an integer number of seconds (if the time is 0.1 it outputs 1, if it is 13.4 it outputs 14), compute the pmf of the reading from the device. What kind of random variable is this?

The pdf of this exponential random variable will be

$$F_{\tilde{t}}(\tilde{t} > 0)(t) = \int_{(x-1)}^x \lambda e^{-\lambda t} dx = e^{-\lambda x-1} - e^{-\lambda x}$$

Which is equal to

$$= (e^{-\lambda})^{x-1} \times (1 - e^{-\lambda}) \text{ when } x \in \mathbf{Z} \text{ when } x \geq 1$$

- (b) What is the pdf of the error between your reading and the true time of decay?

We know that when the conditional cdf of \tilde{t} given $\tilde{t} \geq t_0$ evaluated at $t > t_0$ is given by:

$$F_{\tilde{t}|\tilde{t} \geq t_0} = P(\tilde{t} \leq t | \tilde{t} \geq t_0) = \frac{e^{-\lambda} \times (e^{\lambda t-1})}{1 - e^{-\lambda}} = \frac{e^{\lambda x} - 1}{e^{\lambda} - 1}$$

And if we differentiate the cdf we have obtain we get:

$$\frac{e^{\lambda t} - 1}{e^{\lambda} - 1} dt = \frac{\lambda e^{\lambda t}}{e^{\lambda} - 1}$$

Of course, cdfs are cumulative probability functions, and only hold inbetween the values of 0 and 1, so therefore the function is only valid for that range. That is to say:

$$f_t(t) = \begin{cases} \frac{\lambda e^{\lambda t}}{e^{\lambda}-1}, & 0 \leq t \leq 1 \\ 0, & \text{in any other case} \end{cases}$$

3. (Triangular pdf) We are interested in fitting a model with a parametric pdf equal to

$$f_w(x) = \begin{cases} \frac{2x}{w^2}, & \text{for } 0 \leq x \leq w, \\ 0, & \text{otherwise,} \end{cases} \quad (3)$$

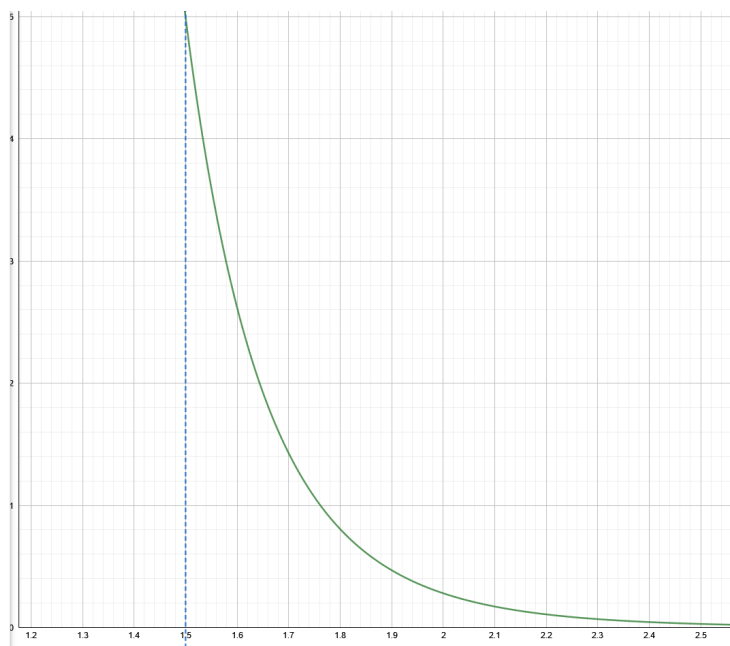
where the parameter w is nonnegative.

- (a) The observed values are 1.25, 0.4, 1.5, 1, 1.2. What are the possible values of the parameter w ?

$w \geq 1.5$ is the current constraint, as we fit a model to our empirical data using the maximum value of the datapoints that we observed, which is equal to (1.5). Therefore $w \geq 1.5$

- (b) Compute the likelihood function corresponding to these data and sketch it.

$$\prod_{i=1}^n \frac{2x_i}{w^2} \rightarrow n = 5 \rightarrow \frac{2^5}{w^{10}} \prod_{i=1}^n x_i = \frac{28.8}{w^{10}} \rightarrow \text{W is maximized at } w = 1.5$$



- (c) What is the maximum likelihood estimate of w ?

Looking at the graph we can see that the maximum likelihood is $MLE = 1.5$ as our pdf is always decreasing and begins at value $w = 1.5$.

- (d) If we observe 100 independent samples that are generated according to the parametric model with a fixed value of w , do you think that there is any chance that the ML estimate of w is correct? Justify your answer intuitively.

No, its intuitive to understand that since we generated this MLE with such a small sample size, and fixed to a maximum of 1.5 as that was our MLE, it will not be representative of a other actual data points as the probability of getting those values according to our MLE will be 0%.

- (e) Let us assume $w = 2$. Generate a sample from a random variable following the model using a uniform sample from the interval $[0, 1]$ equal to 0.64.

$$\text{Likelihood function} \rightarrow pdf \rightarrow y = \frac{2x}{w^2} = \frac{2x}{4} \rightarrow cdf = \int_0^1 \frac{2x}{4} = \frac{x^2}{4}$$

$$\text{Get the inverse cdf } x = \frac{y^2}{4} \rightarrow y = \sqrt{4x} \text{ Evaluate from interval 0 to .64} \rightarrow \sqrt{4(.64)} - 0 = 1.6$$

4. (Applying the cdf) The array in `samples.npy` contains 1,000 i.i.d. samples from an exponential distribution with parameter $\lambda := 1$. Let F denote the cdf of this distribution.

- (a) Apply F to the data (i.e. for each data point x compute $F(x)$) and plot the corresponding histogram. What shape does it have?

It has the shape of an approximate uniform distribution. See below:

```

In [28]: import numpy as np
import math
import os
import matplotlib as plt
import pandas as pd
os.chdir('C:\\Users\\Giulio\\OneDrive\\Desktop')

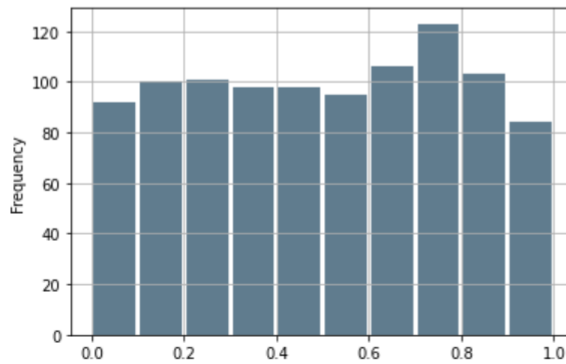
#Import our data
data = np.load('samples.npy')

#Alter our data to create a cdf of a exponential distribution
data_cdf = lambda x: 1-math.exp(x*-1)
cdf_func = np.vectorize(data_cdf, otypes=[float])
transformed_data = cdf_func(data)

#Plot our data
series_data = pd.Series(transformed_data)
series_data.plot.hist(grid=True, bins=10, rwidth = 0.9, color='#607c8e')

```

Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x194f43c5408>



- (b) Let \tilde{a} be a random variable with an invertible cdf $F_{\tilde{a}}$. What is the distribution of $F_{\tilde{a}}(\tilde{a})$? Justify your answer mathematically.

We know that if a cdf is invertible it represents the cdf a continuous random variable, and due to the universality of the norm if we plug in the random variable \tilde{a} into its own cdf $F_{\tilde{a}}(\tilde{a})$ then the distribution that follows will map to the points $[0, 1]$ and be a uniform distribution. We can prove this property in the following way.

Since $F_{\tilde{a}}(\tilde{a})$ takes in values $(0, 1)$, $P(F_{\tilde{a}}(\tilde{a}) \leq x) = 0$ for $x < 0$ and equals 1 for $x \geq 1$. For all $x \in (0, 1)$:

$$P(F_{\tilde{a}}(\tilde{a}) \leq x) = P(\tilde{a} \leq F_{\tilde{a}}^{-1}(x)) = F_{\tilde{a}}(F_{\tilde{a}}^{-1}(x)) = x$$

Therefore $F_{\tilde{a}}(\tilde{a})$ has the univeform distribution $(0, 1)$ CDF.