ABA Recitation 1

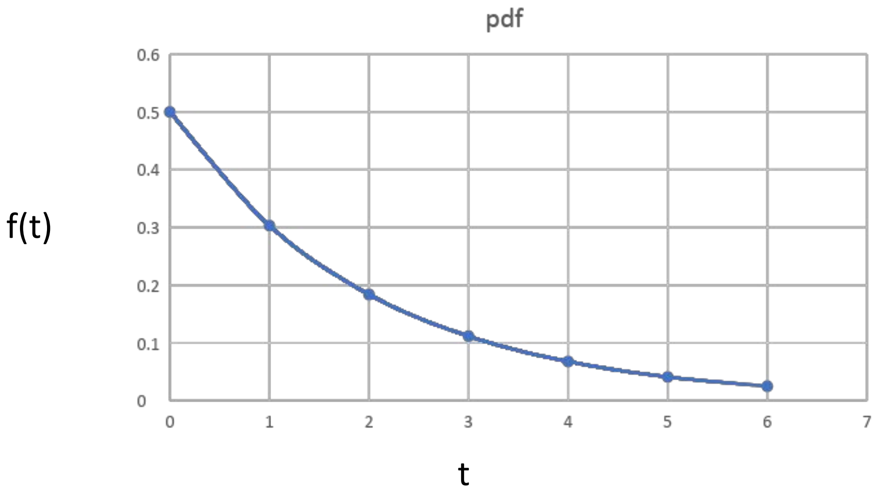
Spring 2024
01/26/24

Agenda

- Quiz 1 is out, due Jan 29 11:59PM
- Hazard Functions:
  - PDF and CDF
  - Hazard function example (constant Hazard)
  - From hazard to CDF F(t) and PDF f(t)
  - Empirical Hazard function (google sheet)
- OLS example in Python

Probability Density Function (PDF) and CDF (Cumulative Density Function)

- What is pdf?
- For Exponential pdf f(t) = $\lambda e^{-\lambda t}$
- f(t) outlines how data is distributed with time t, given the parameter λ. The following figure shows how data is distributed.



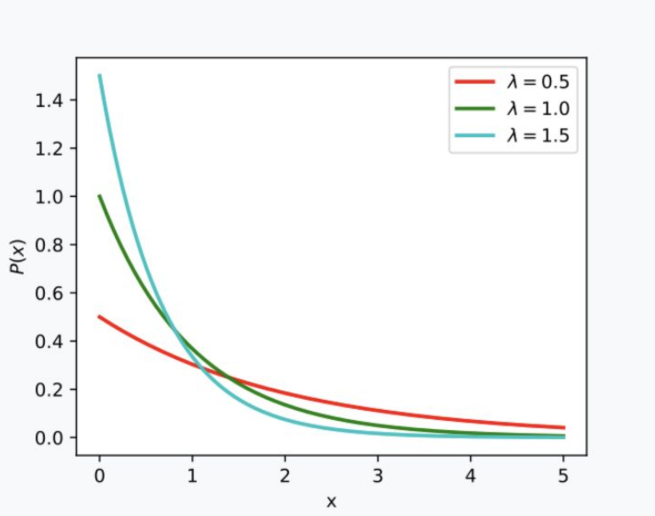
F(t)

- Cumulative distribution is
  - $F(t) = \int_0^t f(x)dx$
  - $F(t) = \int_0^t \lambda e^{-\lambda x} dx$

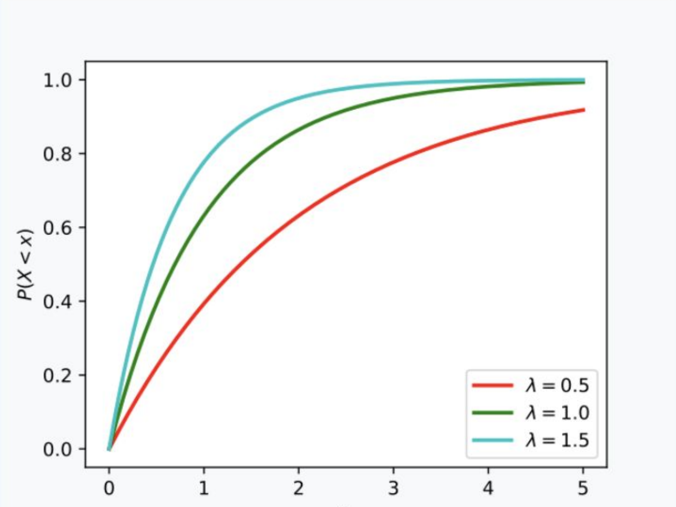
  - Or $F(t) = \lambda \int_0^t e^{-\lambda x} dx$
  - Or $\lambda \ (-\frac{1}{\lambda} [e^{-\lambda x}]_0^t$
- Or (1- $e^{-\lambda t}$)

Exponential

Probability density function



Cumulative distribution function



Hazard Functions

- For analysing **timing question** (when will something happen?)
- Hazard:
  - What is the probability that the event will happen *right now*, given that it has *not yet happened?*
    - This is called the *hazard rate*.

Starting Point for Hazard Analysis

- Exponential function is a commonly used function.

$$
\begin{aligned}
h(t) &= \frac{f(t)}{1 - F(t)} \\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\
&= \lambda
\end{aligned}
$$

- Notice that exponential implies a "constant" hazard rate.
  - This is why exponential is called a "memory-less" distribution because hazard is not a function of time.
- S(T) = 1-F(T) is the survivor function.

Hazard Rate and Distribution Function

- Remember that Hazard functions have one-on-one mapping with the distribution function. If you are defining one, you are automatically defining the other.

$$F(t) = 1 - \exp\left(-\int_0^t h(u)\,du\right)$$

  - Sometimes it is much easier to define Hazard function first.
  - By defining hazard rate, one has timing process to work with.

Derivation: from h(t) to F(t) to f(t) for exponential

- In exponential hazard = λ

Derivation: from h(t) to F(t) to f(t) for exponential

- In exponential hazard = λ
- We know that

$$F(t) = 1 - \exp\left(-\int_0^t h(u)\, du\right)$$

- Since h() is a constant, integration of constant is simply h(). u or

So $F(t) = 1 - \exp(-\int_0^t \lambda\, du)$

Or $F(t) = 1 - \exp(-\lambda t)$

Distributions

- Commonly used distribution for survival analysis?

  Exponential pdf

  $$f(t) = \lambda e^{-\lambda t}$$

  Exponential cdf

  $$F(t) = 1 - e^{-\lambda t}$$

  Hazard

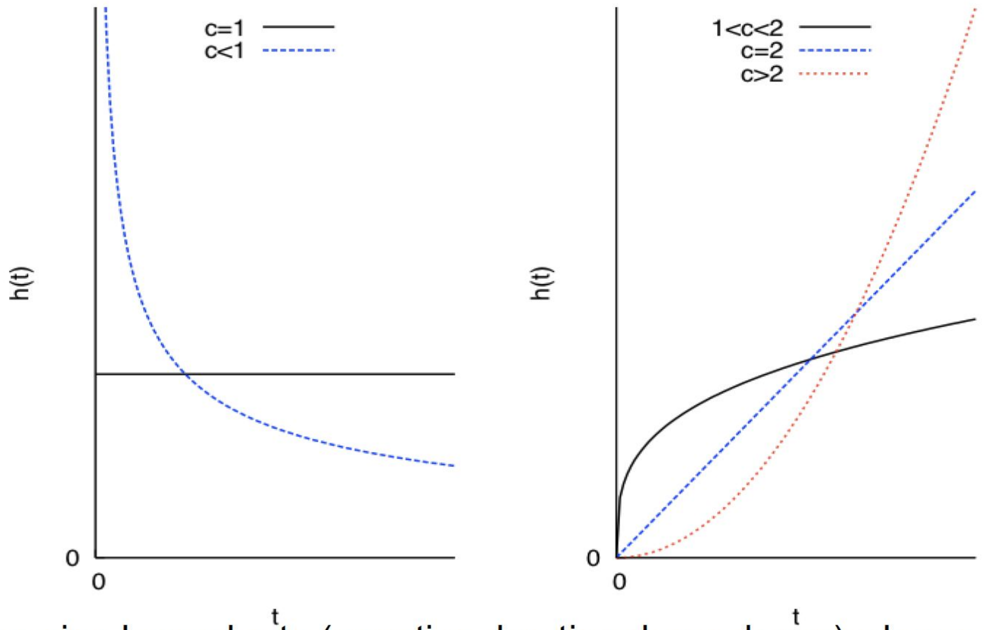
  $$h(t) = \frac{f(t)}{1 - F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

  Constant hazard. That is the hazard rate is not a function of time.

Weibull Distribution

- A flexible distribution that can represent increasing or decreasing hazard rates readily. This allows for λ to change with time
  - $F(t) = 1 - e^{-\lambda t^c}$
  - $f(t) = \lambda c t^{c-1} e^{-\lambda t^c}$
  - $h(t) = \frac{f(t)}{1-F(t)} = \lambda c t^{c-1}$
- The hazard is function of time. That means depending on c, the hazard can be increasing or decreasing. For c=1, this boils down to exponential

Weibull Hazard



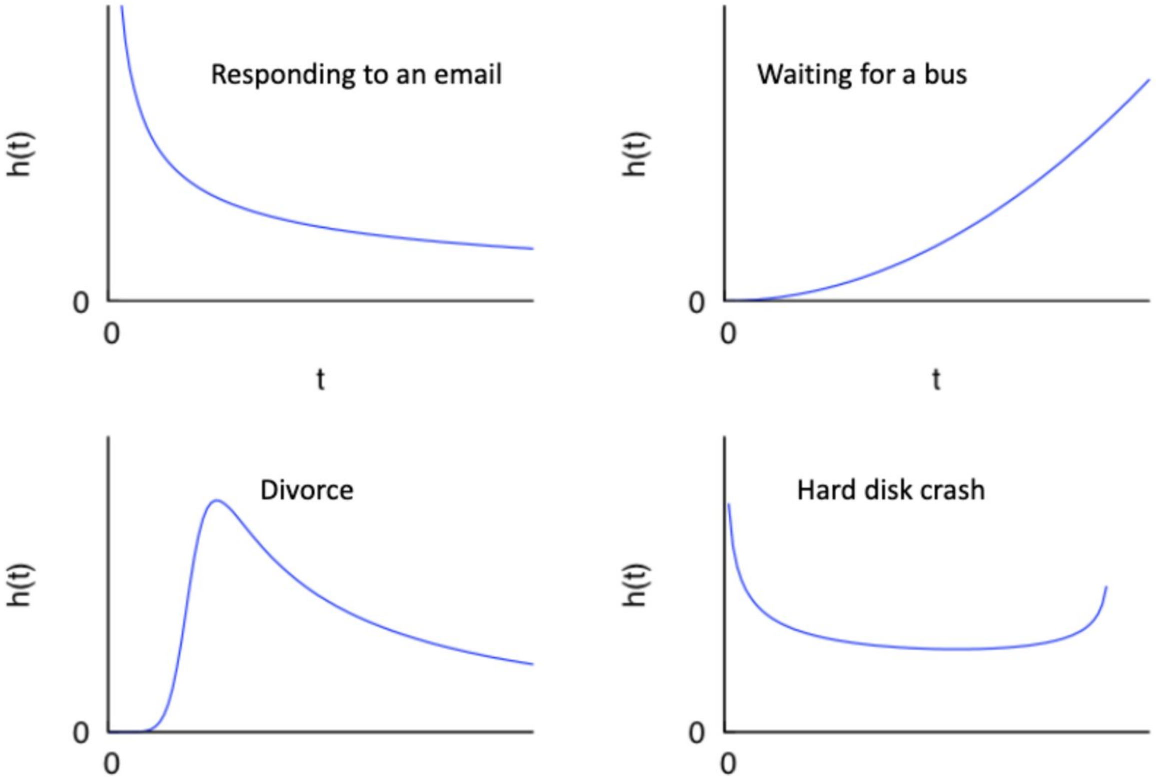
- Decreasing hazard rate (negative duration dependence) when c<1

- Increasing hazard rate (positive duration dependence) when c>1

Empirical Hazard Example

See google sheet:
https://docs.google.com/spreadsheets/d/1feqc_Sry5nKtmHXAkZkFGTA3TRANnq9GNJMmOV0aS8k/edit?usp=sharing

Some shapes of hazard rate functions



OLS in Python

See jupyter notebook

Note on libraries for next homeworks

install lifelines to use for KM in python

```
#conda install -c conda-forge lifelines
```