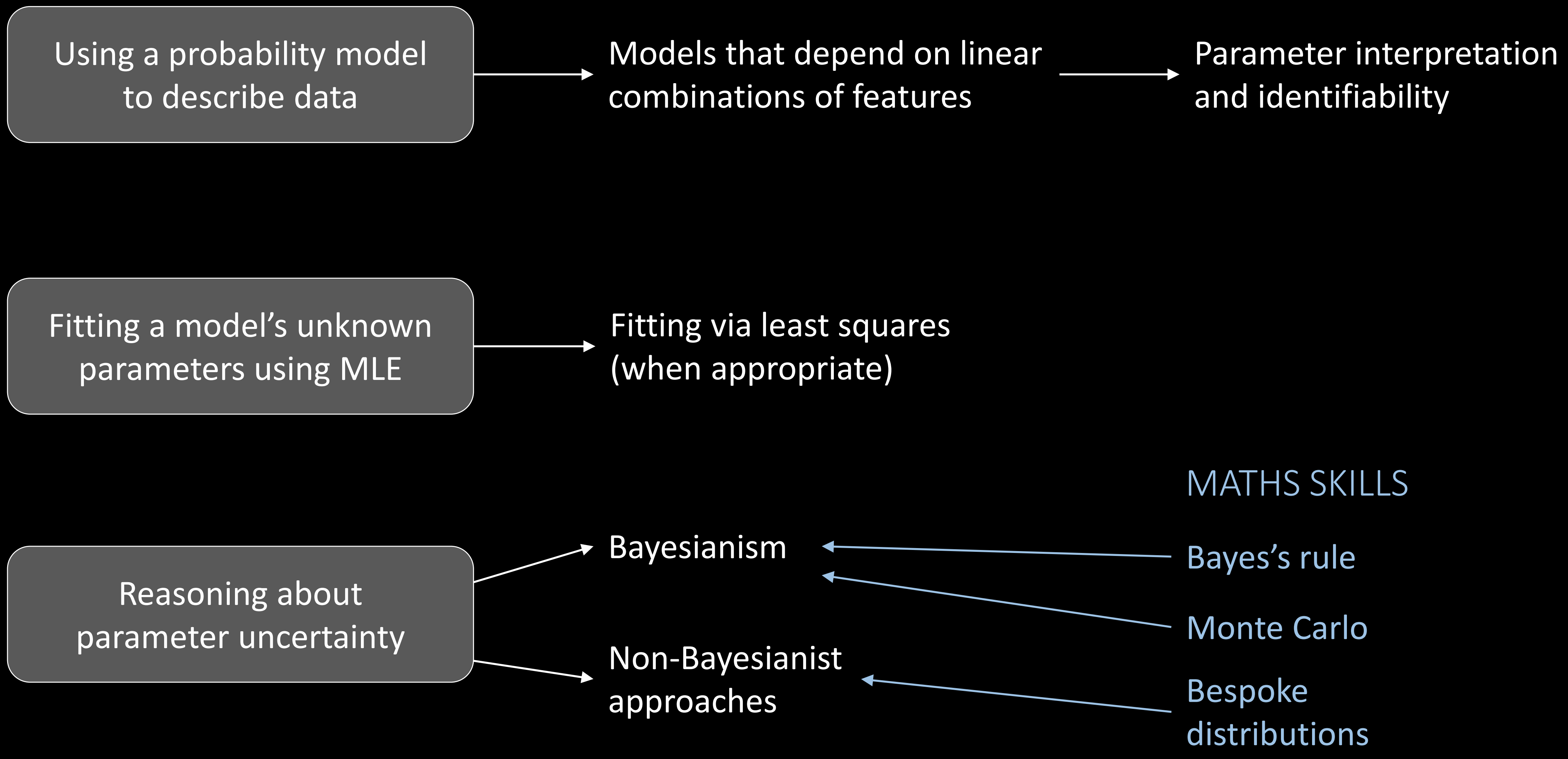
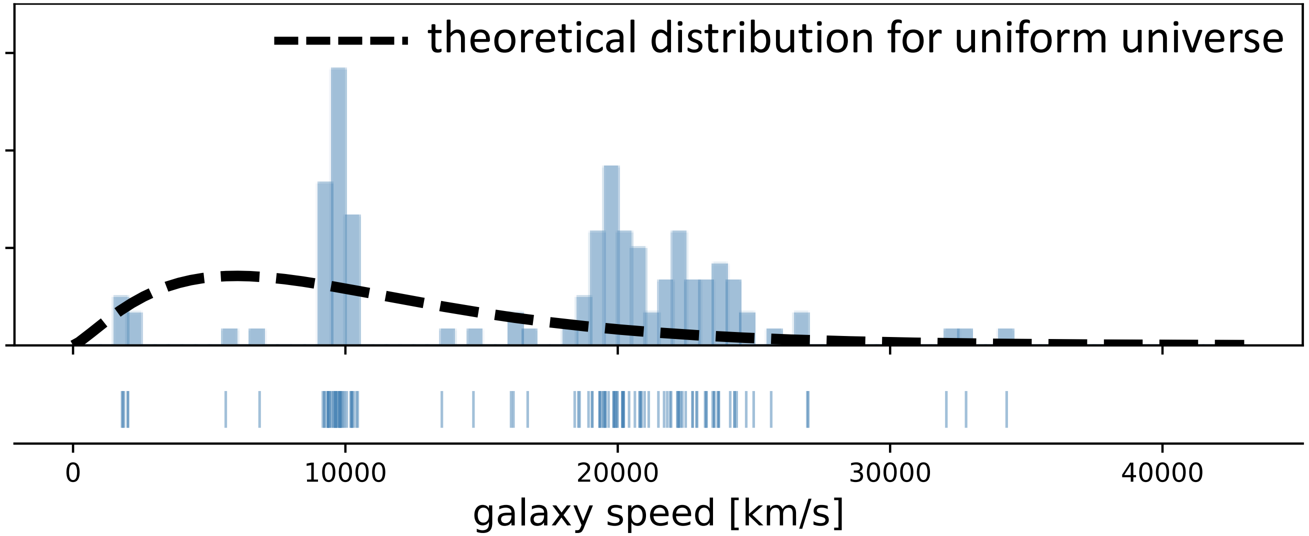
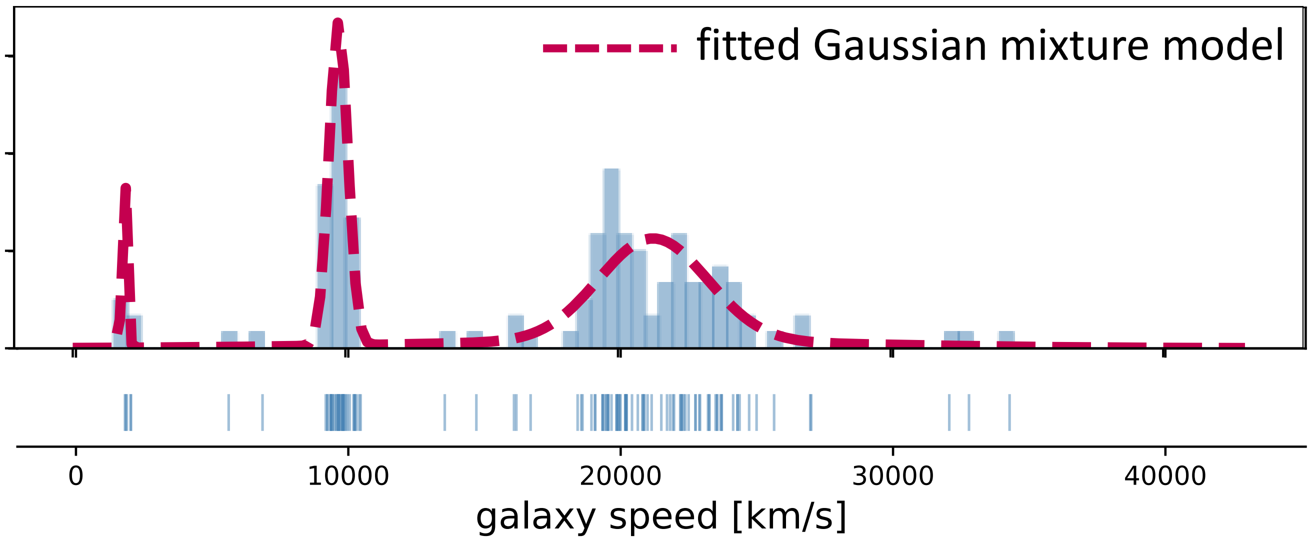
IB Data Science syllabus



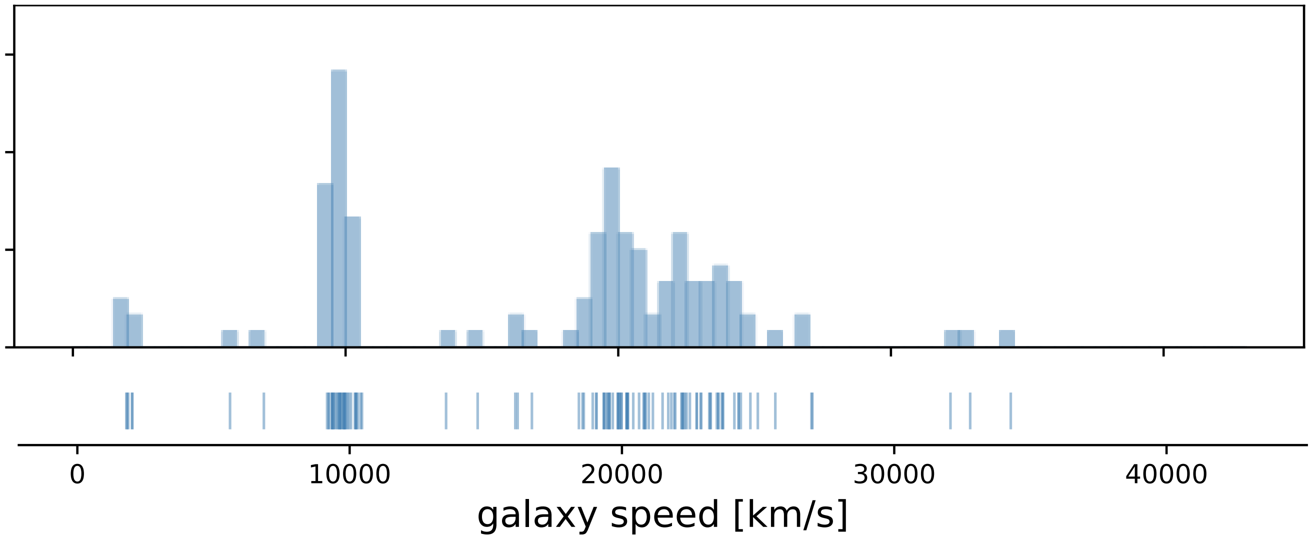
This chart shows the distribution of the speeds of 120 galaxies, from a survey of the Corona Borealis region.
Postman, Huchra, Geller (1986)





fitted Gaussian mixture model
0
10000
20000
30000
40000
galaxy speed [km/s]

What's the best distribution we can find, to model this dataset?



There are four ways to specify a distribution.

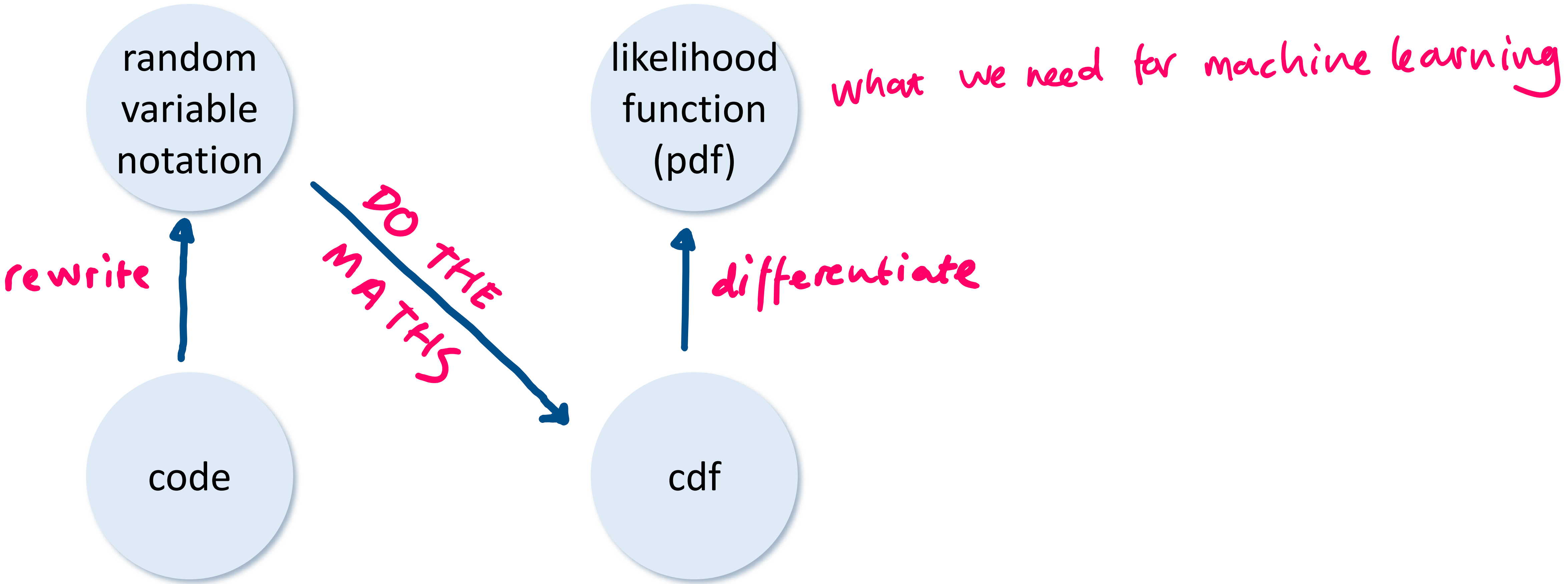
random variable notation

likelihood function (pdf)

code

cdf

Bespoke probability distributions part I: from code to likelihood (for continuous random variables)



Exercise 5.3.2

Find the pdf of the random variable generated by this code:

```
u = np.random.uniform()
x = - np.log(u) / λ        # λ>0
```

STRATEGY

Try to write our probability in terms of simple standard random variable (for which we can look up the cdf)

Break it down so that the random variables are on the left (so we can use the cdf)

Step 1: random variable notation

$U \sim U[0,1]$

$X = -\frac{1}{\lambda}\log U$

Step 2: X is a continuous r.v., so find its cdf

cdf for X: $\mathbb{P}(X \leq x) = \mathbb{P}\left(-\frac{1}{\lambda}\log U \leq x\right)$ by defn. of X

$= \mathbb{P}(U \geq e^{-\lambda x})$

$= 1 - \mathbb{P}(U < e^{-\lambda x})$

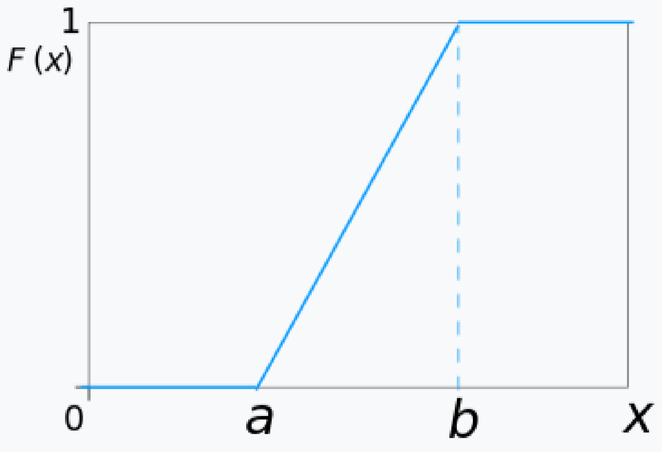
$= \begin{cases} \text{if } x \geq 0, \; e^{-\lambda x} \leq 1, \text{ so we get: } 1 - \left(\frac{e^{-\lambda x} - 0}{1 - 0}\right) = 1 - e^{-\lambda x} \\ \text{if } x < 0: \ldots \end{cases}$

Step 3: differentiate cdf to get pdf

$\frac{d}{dx}\, 1 - e^{-\lambda x} = \lambda e^{-\lambda x}.$

Wikipedia: Uniform distribution

Cumulative distribution function



Notation	$\mathcal{U}_{[a,b]}$
Parameters	$-\infty < a < b < \infty$
Support	$x \in [a, b]$
PDF	$\begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$
CDF	$\begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } x \in [a,b] \\ 1 & \text{for } x > b \end{cases}$

Exercise 5.3.3

Find the density of the random variable generated by this code:

```
def ry():
    x = np.random.uniform()
    return x **2
```

Exercise 5.3.5 (Gaussian mixture model)

Find the likelihood function for the Gaussian mixture model.

Exercise.

(See lecture notes for solution.)