



Day 10: Pandas: Series & Data Frames; Probability & Statistics

Paul Schumacher, MSc Quantitative Economics

Announcements



Attendance

Final Review

Thursday 4pm – 5:30pm

Class Survey
<https://forms.gle/QnCN7MnSHD5dC1Xz9>

HW2 due
Tonight 11pm

Final
Friday 4pm – 6pm

Pandas: Series & Data Frames

Series: “1 Column in Data Set”

```
s = pd.Series(  
    np.random.randn(5),  
    index=['a', 'b', 'c', 'd', 'e'],  
    name='example')
```

```
s  
  
a    -1.066690  
b    -0.640083  
c     0.253090  
d     0.615389  
e    -0.053631  
Name: example, dtype: float64
```

“n Columns in Data Set”

Data Frames:

```
d = {'one' : pd.Series([1., 2., 3.], index=['a', 'b', 'c']),  
     'two' : pd.Series([1., 2., 3., 4.], index=['a', 'b', 'c', 'd'])}
```

```
df = pd.DataFrame(d)
```

```
df
```

	one	two
a	1.0	1.0
b	2.0	2.0
c	3.0	3.0
d	NaN	4.0

Exam Prep

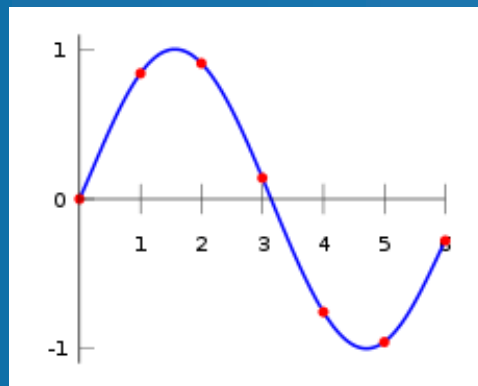
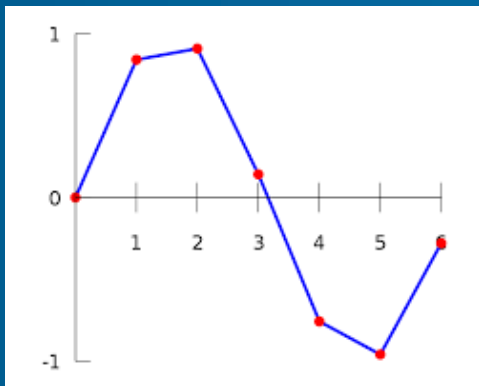


- “PLF Day 10 Exam Prep.ipynb”
- Breakout rooms
- Time: 15 - 20 min

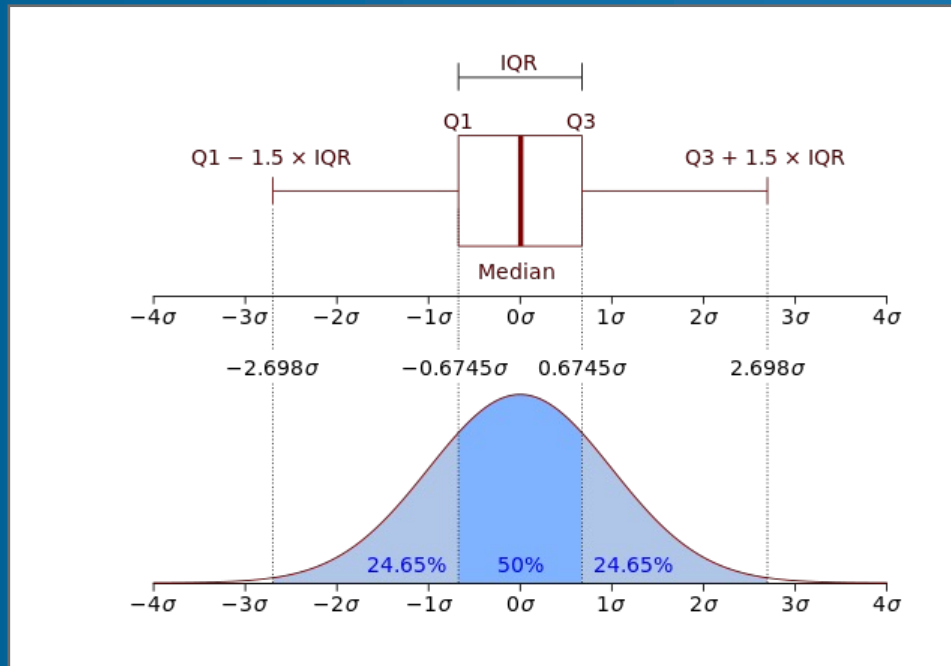
Probability & Statistics: SciPy



- “Advanced NumPy” → Scientific Python
 - Optimization, Linear Algebra, Advanced Statistics, Signal Processing
 - Interpolation (finding new data points):

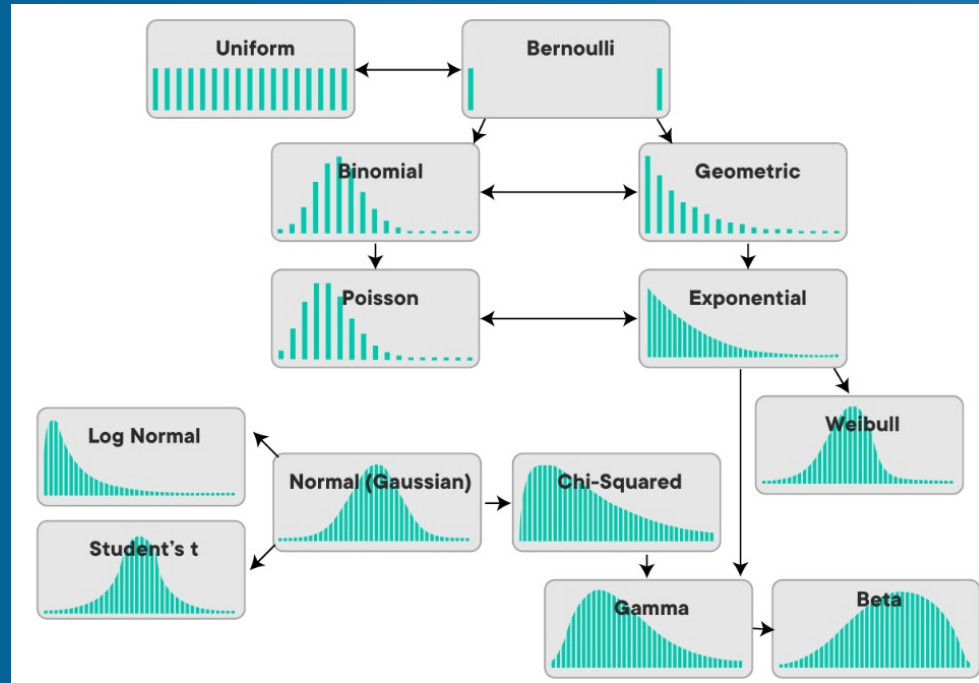


Probability & Statistics: PDF



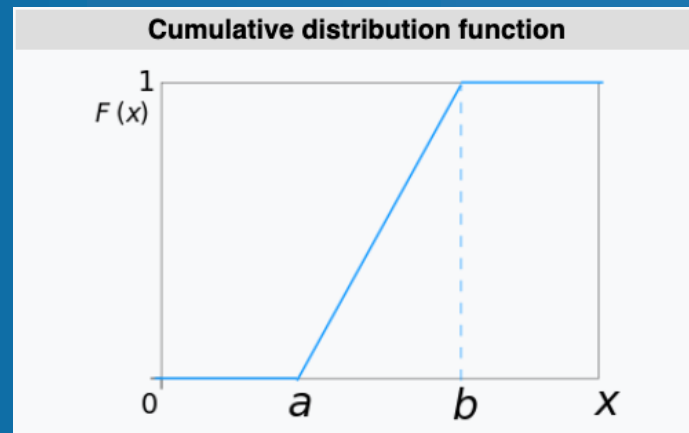
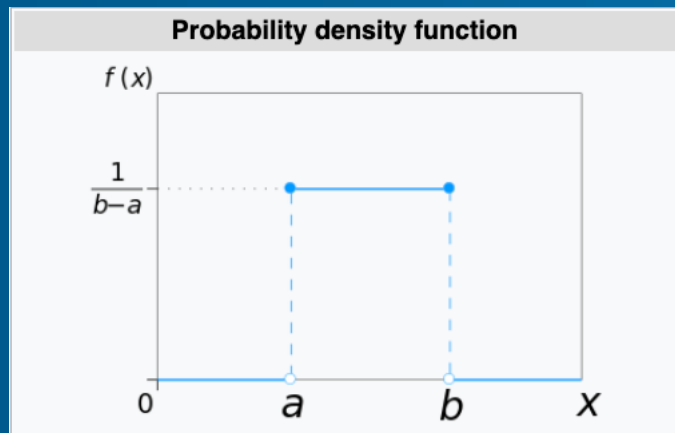
Probability
Density
Function

Distribution Types



Probability & Statistics: CDF

Cumulative Distribution Function





Any Questions?