stellar_mass_distribution_nonparametric_test_kolmogorov_smirnov

March 12, 2021

1 KEPLER Exoplanets Database

1.1 Star mass distribution for stars with exoplanets

Source: https://data.world/markmarkoh/kepler-confirmed-planets/workspace/project-summary?agentid=markmarkoh&datasetid=kepler-confirmed-planets NASA Exoplanet archive: https://exoplanetarchive.ipac.caltech.edu/docs/data.html

@Author: Javier Cebrián

@Attention: In this file there are Plotly (rendered with HTML) plots. If you are viewing it with github, please enable external view with nbviewer

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import plotly.offline as pyo
from scipy import stats
import plotly.io as pio
pio.renderers.default = "notebook+pdf"
```

[2]:
```
planetsDf=pd.read_csv('../planets.csv',delimiter=',')
```

[3]:
```
planetsDf
```

[3]:
```
      rowid pl_hostname pl_letter    pl_discmethod  pl_pnum    pl_orbper  \
0         1      11 Com         b  Radial Velocity        1   326.030000
1         2      11 UMi         b  Radial Velocity        1   516.220000
2         3      14 And         b  Radial Velocity        1   185.840000
3         4      14 Her         b  Radial Velocity        1  1773.400000
4         5    16 Cyg B         b  Radial Velocity        1   798.500000
…       …           …        …                …        …           …
3367   3368     ups And         b  Radial Velocity        4     4.617033
3368   3369     ups And         c  Radial Velocity        4   241.258000
3369   3370     ups And         d  Radial Velocity        4  1276.460000
```

```
3370   3371     ups And         e  Radial Velocity        4  3848.860000
3371   3372      xi Aql         b  Radial Velocity        1   136.750000

      pl_orbpererr1  pl_orbpererr2  pl_orbperlim  pl_orbsmax  …  \
0          0.320000      -0.320000           0.0    1.290000  …
1          3.250000      -3.250000           0.0    1.540000  …
2          0.230000      -0.230000           0.0    0.830000  …
3          2.500000      -2.500000           0.0    2.770000  …
4          1.000000      -1.000000           0.0    1.681000  …
…             …            …           …         …  …
3367       0.000023      -0.000023           0.0    0.059222  …
3368       0.064000      -0.064000           0.0    0.827774  …
3369       0.570000      -0.570000           0.0    2.513290  …
3370       0.740000      -0.740000           0.0    5.245580  …
3371       0.250000      -0.250000           0.0    0.680000  …

      st_masserr1  st_masserr2  st_masslim  st_massblend  st_rad  st_raderr1  \
0             0.30        -0.30         0.0           0.0   19.00        2.00
1             0.25        -0.25         0.0           0.0   24.08        1.84
2             0.10        -0.20         0.0           0.0   11.00        1.00
3             0.05        -0.05         0.0           0.0     NaN         NaN
4              NaN          NaN         0.0           0.0     NaN         NaN
…           …          …         …           …     …         …
3367           NaN          NaN         0.0           0.0    1.56         NaN
3368           NaN          NaN         0.0           0.0    1.56         NaN
3369           NaN          NaN         0.0           0.0    1.56         NaN
3370           NaN          NaN         0.0           0.0    1.56         NaN
3371           NaN          NaN         0.0           0.0   12.00         NaN

      st_raderr2  st_radlim  st_radblend   rowupdate
0          -2.00        0.0          0.0  2014-05-14
1          -1.84        0.0          0.0  2014-05-14
2          -1.00        0.0          0.0  2014-05-14
3            NaN        NaN          0.0  2014-05-14
4            NaN        NaN          0.0  2015-09-10
…          …        …          …         …
3367         NaN        0.0          0.0  2014-05-14
3368         NaN        0.0          0.0  2014-05-14
3369         NaN        0.0          0.0  2014-05-14
3370         NaN        0.0          0.0  2014-05-14
3371         NaN        0.0          0.0  2014-05-14

[3372 rows x 67 columns]
```

There are a lot of stars with two or more planets. In order to count only one time each star I erase all rows from repated stars.

[4]:
```
planetsDf=planetsDf.set_index("pl_hostname")
planetsDf = planetsDf[~planetsDf.index.duplicated(keep='first')]
```

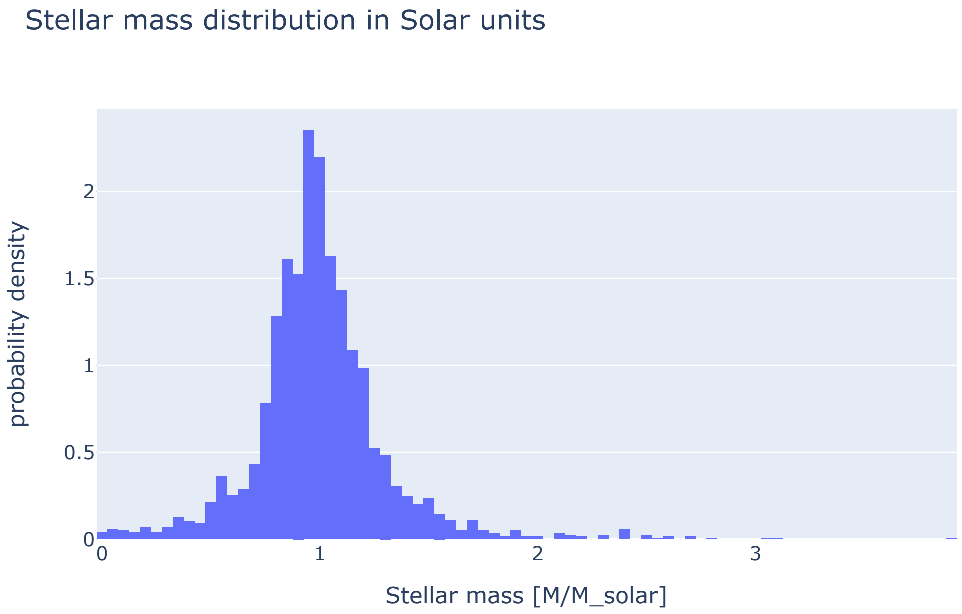
[5]:
```
planetsDf
```

[5]:
```
             rowid pl_letter    pl_discmethod  pl_pnum     pl_orbper  \
pl_hostname
11 Com           1         b  Radial Velocity        1    326.030000
11 UMi           2         b  Radial Velocity        1    516.220000
14 And           3         b  Radial Velocity        1    185.840000
14 Her           4         b  Radial Velocity        1   1773.400000
16 Cyg B         5         b  Radial Velocity        1    798.500000
…              …       …                …      …           …
psi Dra B     3365         b  Radial Velocity        1   3117.000000
tau Boo       3366         b  Radial Velocity        1      3.312457
tau Gem       3367         b  Radial Velocity        1    305.500000
ups And       3368         b  Radial Velocity        4      4.617033
xi Aql        3372         b  Radial Velocity        1    136.750000

             pl_orbpererr1  pl_orbpererr2  pl_orbperlim  pl_orbsmax  \
pl_hostname
11 Com            0.320000      -0.320000           0.0    1.290000
11 UMi            3.250000      -3.250000           0.0    1.540000
14 And            0.230000      -0.230000           0.0    0.830000
14 Her            2.500000      -2.500000           0.0    2.770000
16 Cyg B          1.000000      -1.000000           0.0    1.681000
…                        …              …             …           …
psi Dra B        42.000000     -42.000000           0.0    4.430000
tau Boo           0.000007      -0.000007           0.0    0.049000
tau Gem           0.100000      -0.100000           0.0    1.170000
ups And           0.000023      -0.000023           0.0    0.059222
xi Aql            0.250000      -0.250000           0.0    0.680000

             pl_orbsmaxerr1  …  st_masserr1  st_masserr2  st_masslim  \
pl_hostname                  …
11 Com                0.050  …         0.30        -0.30         0.0
11 UMi                0.070  …         0.25        -0.25         0.0
14 And                  NaN  …         0.10        -0.20         0.0
14 Her                0.050  …         0.05        -0.05         0.0
16 Cyg B              0.097  …          NaN          NaN         0.0
…                         …  …            …            …           …
psi Dra B             0.040  …         0.07        -0.07         0.0
tau Boo               0.003  …         0.05        -0.05         0.0
tau Gem                 NaN  …         0.30        -0.30         0.0
ups And               0.000  …          NaN          NaN         0.0
xi Aql                  NaN  …          NaN          NaN         0.0
```

```
              st_massblend  st_rad  st_raderr1  st_raderr2  st_radlim  \
pl_hostname
11 Com                 0.0   19.00        2.00       -2.00        0.0
11 UMi                 0.0   24.08        1.84       -1.84        0.0
14 And                 0.0   11.00        1.00       -1.00        0.0
14 Her                 0.0     NaN         NaN         NaN        NaN
16 Cyg B               0.0     NaN         NaN         NaN        NaN
…                        …       …           …           …          …
psi Dra B              0.0     NaN         NaN         NaN        NaN
tau Boo                0.0    1.46        0.05       -0.05        0.0
tau Gem                0.0   26.80        0.70       -0.70        0.0
ups And                0.0    1.56         NaN         NaN        0.0
xi Aql                 0.0   12.00         NaN         NaN        0.0

              st_radblend   rowupdate
pl_hostname
11 Com                0.0  2014-05-14
11 UMi                0.0  2014-05-14
14 And                0.0  2014-05-14
14 Her                0.0  2014-05-14
16 Cyg B              0.0  2015-09-10
…                       …           …
psi Dra B             0.0  2015-12-17
tau Boo               0.0  2015-04-16
tau Gem               0.0  2014-05-14
ups And               0.0  2014-05-14
xi Aql                0.0  2014-05-14

[2509 rows x 66 columns]
```

Now probability density can be plotted:

[6]:
```
fig = px.histogram(planetsDf, x="st_mass", histnorm='probability␣
 ↪density',title='Stellar mass distribution in Solar units',width=800,␣
 ↪height=320)
fig.update_xaxes(title='Stellar mass [M/M_solar]')
fig.show()
```



1.2 Although in sight it is not a Gaussian distribution, an interesting exercise is to check it.

1.2.1 - First with a nonparametric test. (i.e. Kolmogorov-Smirnov)

1.2.2 - Second by visual inspection.

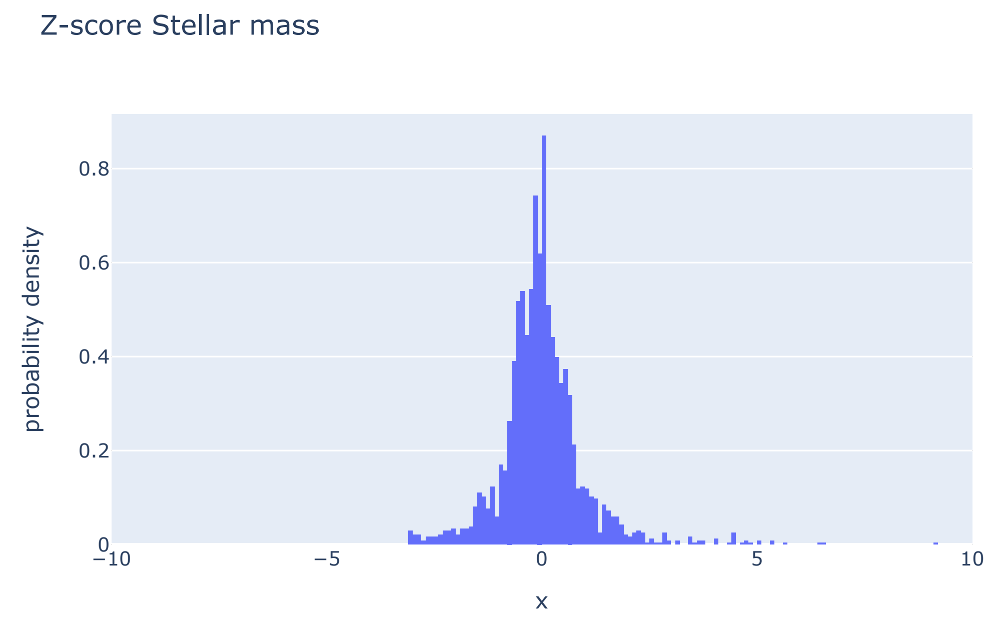
Anyway, before I apply Z-score in order to compare with normal distribution N(0,1)

[7]:

```
zscoreMass=stats.zscore(planetsDf['st_mass'].dropna().
 ↪to_numpy(),nan_policy='omit')
```

[8]:

```
fig = px.histogram( x=zscoreMass, histnorm='probability density',title='Z-score␣
 ↪Stellar mass')
fig.update_xaxes(range=(-10,10))
fig.show()
```



1.3 1- Kolmogorov-Smirnov test

This is a nonparametrical test that compares the distance between the empirical distribution of the sample data with with a reference probability distribution, in this case the normal distribution.

Null-hypothesis = Distributions are equal

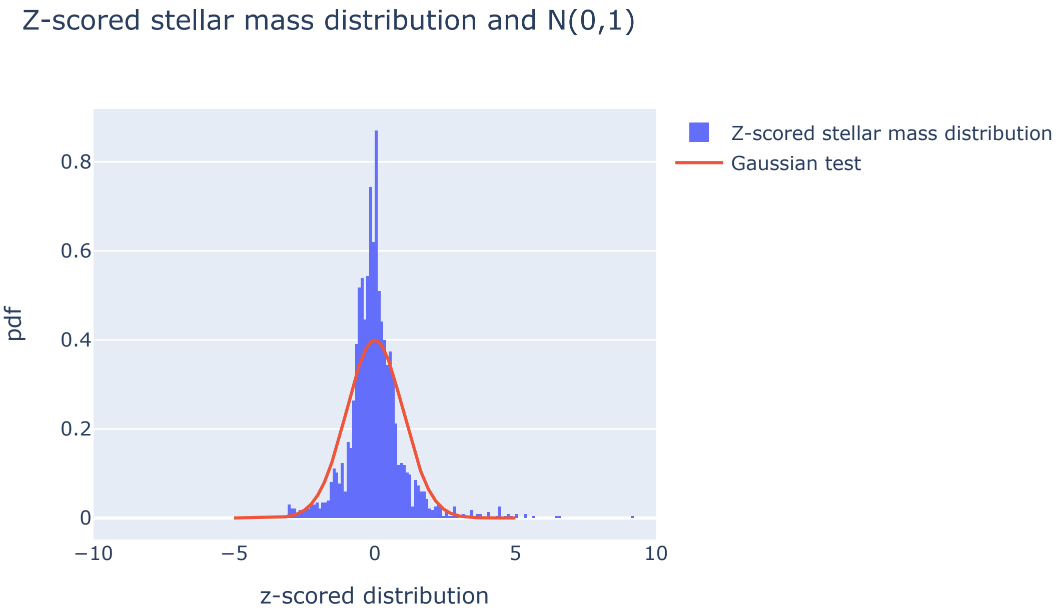
```
[9]: stats.kstest(zscoreMass, 'norm')
```

```
[9]: KstestResult(statistic=0.10757731944365112, pvalue=3.2958363736394215e-24)
```

Test rejects null-hypothesis with a very small p-value. ## 2- Visual inspection

```
[10]: x=np.linspace(-5,5,1000)
      npdf= stats.norm.pdf(x,loc=0,scale=1)

      data0 = go.Histogram(x=zscoreMass, histnorm='probability␣
       ↪density',name='Z-scored stellar mass distribution')
      data1 = go.Scatter( x=x, y=npdf,mode='lines',name='Gaussian test')

      data = [data0, data1]

      layout = go.Layout(title='Z-scored stellar mass distribution and N(0,1)')

      fig = go.Figure(data= data, layout = layout)
```

```
fig.update_xaxes(range=(-10,10), title='z-scored distribution')
fig.update_yaxes(title='pdf')
fig.show()
#pyo.plot(fig, filename = 'line_chart.html')
```



Finally, with visual inspection it is obvious that it does not match.

1.4 Conclusions

According to the data, the most representative star with planets has the same mass than our Sun.