MA675 Astronomy Project

Group 4: Lihao Liao, Yingjie Wang, Yuchen Liu, Ruining Jia

Background

A Stellar occultation is when a solar system body briefly blocks a distant planet, which results in a gradual decrease of the star's light. Observations of an occultation produce a light curve, which can be used to determine the occulting planet's atmosphere properties [from light curve to atmosphere properties]. Reversely, we can also forward model an atmosphere profile to predict the occultation light curve that would result from an occultation by that atmosphere [from atmosphere properties to light curve].

The goal of this project is to compare the forward model to a suite of archival observed light curve and trying to determine if they are consistent or inconsistent. For example, in Figure 1, three forward models are shown compared to observed datasets in blue and black line is a binned version of the blue data.
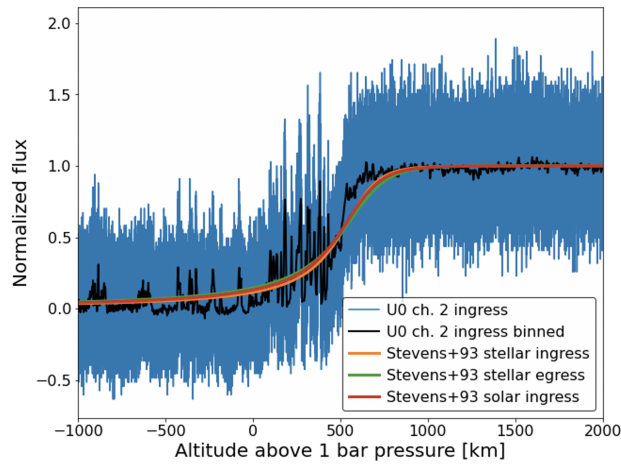

Figure 1: Forward models compared with observed data in blue. Black line is a binned version of the blue data.

Therefore, our null hupothesis is: the observed data is drawn from the same underlying distribution as the forward model results.

About Kolmogorov-smirnov Test

K-S test is a nonparametric statistical test method for continuous distribution:

1.Single sample test

Example code: ks.test(data, distribution, parameter)

Such as : ks.test(x."pgamma",3,2)

2.Double sample test: Example code: ks.test(data1, data2)

$$K = \sup_{t \in [0,1]} |B(t)|$$

Equation:

B(t): Brownian Bridge

Also:

$$\Pr(K \leq x) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{k=1}^{\infty}e^{-(2k-1)^2\pi^2/(8x^2)},$$

Limitations of K-S Test

1. Continuous distribution
2. Nonparametric test method

Kullback-Leibler Divergence & Jensen-Shannon Divergence

K-L Divergence: Measure the distance in probability distribution space, but the dicrimination of similarity is not accurate due to the asymmetry of its structure .

J-S Divergence: Variant of K-L divergence which solves the problem of asymmetry of KL divergence.

J-S Divergence

Equation:

$$D_{JS}(f \| g) = \frac{1}{2}D_{KL}(f \| \frac{f+g}{2}) + \frac{1}{2}D_{KL}(g \| \frac{f+g}{2})|$$

Where:

$$D_{KL}(f \| g) = \int_{-\infty}^{\infty} f(x)\log(\frac{f(x)}{g(x)})dx$$

Discrete: also could be written as:

$$D_{KL}(f \| g) = \sum_{i} f_i \log(\frac{f_i}{g_i})$$

Characteristics: 1. Symmetric.

```
                2. Range: [0,1]
```

Codes of J-S Divergence

Discrete variables:

import numpy as np

import scipy.stats def JS_divergence(p,q)

M=(p+q)/2

return 0.5*scipy.stats.entropy(p,M)+0.5*scipy.stats.entropy(q, M)

print(JS_divergence(data1, data2))

Continuous variables:

from scipy.stats import norm

import pandas as pd

def JS_div(arr1,arr2,num_bins):

max0 = max(np.max(arr1),np.max(arr2))

min0 = min(np.min(arr1),np.min(arr2))

```python
    bins = np.linspace(min0-1e-4, max0-1e-4, num=num_bins)
    PDF1 = pd.cut(arr1,bins).value_counts() / len(arr1)
    PDF2 = pd.cut(arr2,bins).value_counts() / len(arr2)
    return JS_divergence(PDF1.values,PDF2.values)
print(JS_div(data1,data2,num_bins=20))
```