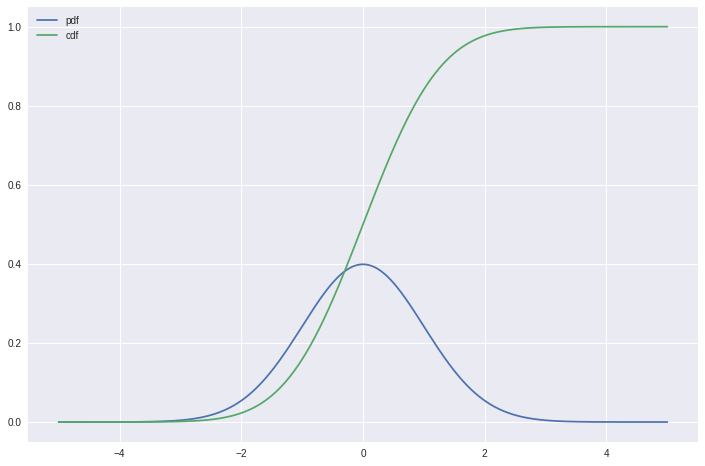
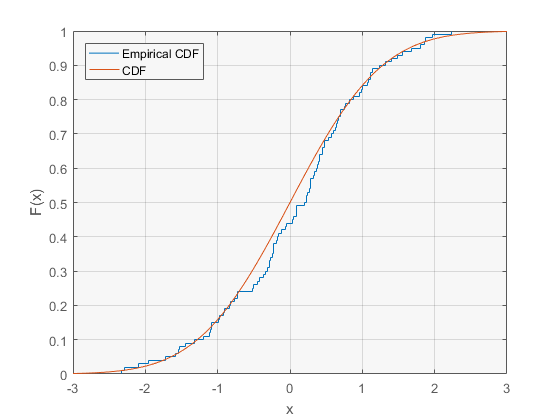
 想請問經驗分布函數(ECDF)的用途為何?

=> 嗨，這邊可以先比較兩種圖：

* 機率密度函數(Probability Density Function, PDF）
* 累積分佈函數(Cumulative Distribution Function CDF)



簡單來說，PDF 是記錄每個值所佔的比例，CDF 就是把 PDF 累加起來的結果。因此 CDF 更可以看出數值間影響關係，例如增加幅度。CDF 是理論值，實際上根據資料畫出來的 CDF 稱為 ECDF。



Reference:

1. [**https://emredjan.github.io/blog/2017/07/19/plotting-distributions/**](https://emredjan.github.io/blog/2017/07/19/plotting-distributions/)

2. [**https://stats.stackexchange.com/questions/239937/empirical-cdf-vs-cdf**](https://stats.stackexchange.com/questions/239937/empirical-cdf-vs-cdf)

**留言 4**



發佈

* **E**

Ella Yang

・2019/05/02 14:37

了解 那這樣資料也應為連續型欄位才能使用吧?

[張維元 (WeiYuan Chang)](https://www.cupoy.com/influencer/A1710885)

・2019/05/02 14:41

離散的話會改用 PMF 取代 PDF，但是 CDF 還是可以用的，只是畫出來的圖會變成階梯狀的。

[張維元 (WeiYuan Chang)](https://www.cupoy.com/influencer/A1710885)

・2019/05/02 14:42

像這樣： https://i.stack.imgur.com/4cyNv.png

查看更多留言

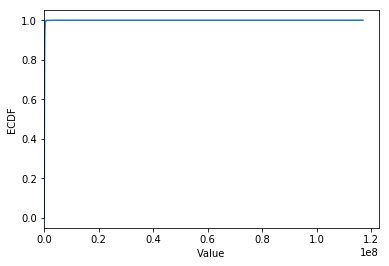
 0

[張維元 (WeiYuan Chang)](https://www.cupoy.com/influencer/A1710885)

・2019/05/02 14:34

是用來判斷異常值嗎?如果是的話 又如何判斷呢?

=> 當 ECDF 畫出來的圖會長這樣：



表示**資料很集中或是有異常值導致資料瞬間累積到極大值**，後者表示有資料異常的可能性

**留言 4**



發佈

* **E**

Ella Yang

・2019/05/02 14:40

了解 那叫才有傳一個stackflow的網址 裡面提到不要用step function，我想請問甚麼又是step function呢?如果不考慮用不用這個function，示不適就能直接用scipy的ECDF套件?

[張維元 (WeiYuan Chang)](https://www.cupoy.com/influencer/A1710885)

・2019/05/02 15:01

可以附一下那個網址嗎？不然我不知道你在講什麼XD

* **E**

Ella Yang

・2019/05/02 16:45

https://stackoverflow.com/questions/14006520/ecdf-in-python-without-step-function