In [34]:

output:

This distribution plot was helpful to tell whether or not the percentage of uppercase in emails is doing anything to seperate spam from ham. You can see that ham (blue)has a high peak and small right tail which means that ham tends to cluster in a predicable way. Spam (green) has a wider distribution, showing that it is seperable from ham using this classifier.

Out[34]:

<matplotlib.axes._subplots.AxesSubplot at 0x113fa2c50>