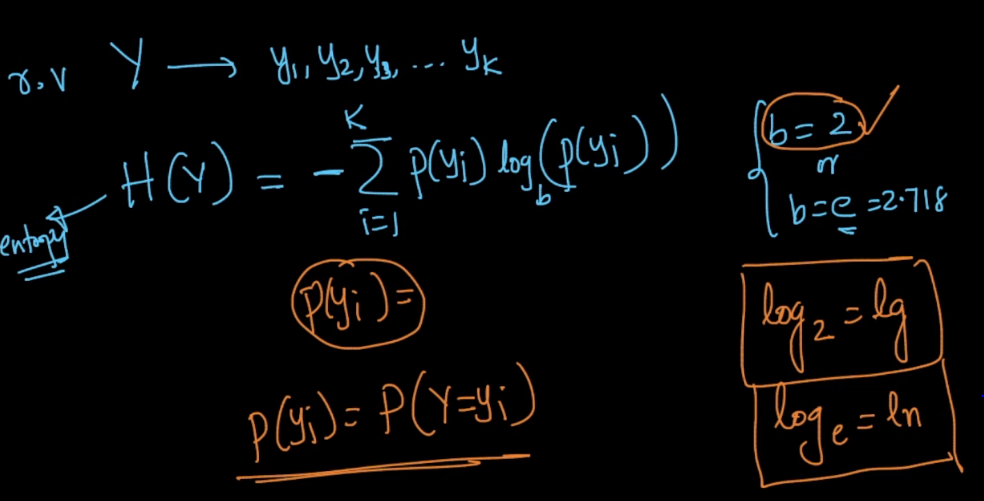
**Entropy:**

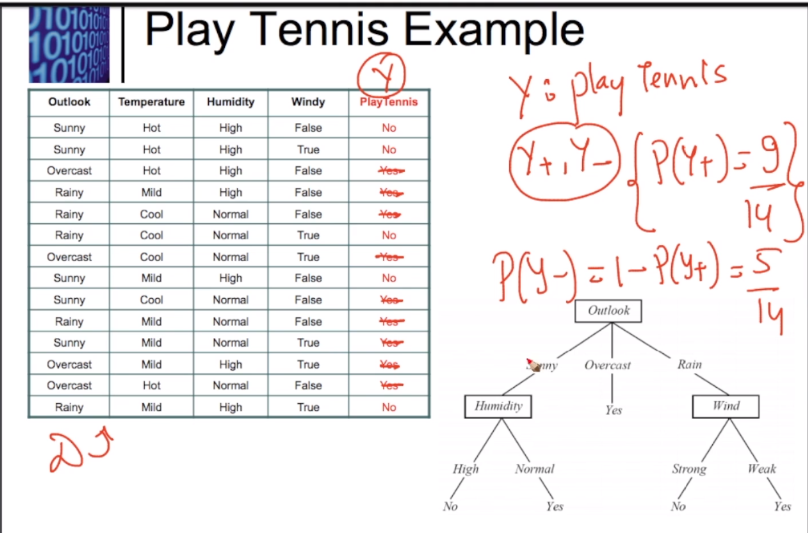
Formula for calculating Entropy is given below.

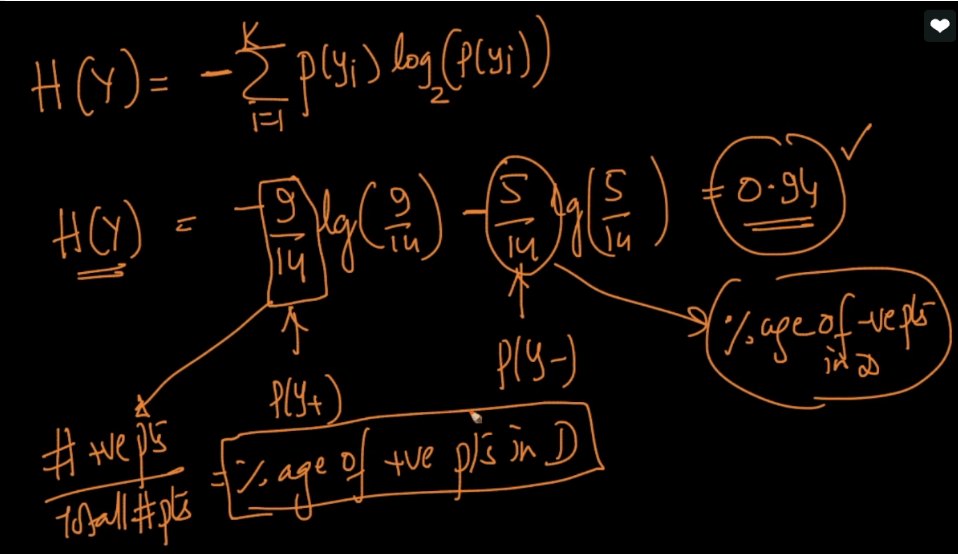
It’s calculated for each feature. Let’s say we calculate it for **Play Tennis**, since it contains 2 types of values **true** or **false.**

So we’ll calculate probability for each type of value ie P(y=true) and P(y=false).

And using this probability we calculate entropy.







Here P(y=true) says what percentage of play tennis feature is **true.**

Now in below figure we can see for 3 different cases:

**1st case:**

If y+ is 99% and y- is 1%, that means 99% value Is true and only 1% is false,

then it’s entropy is 0.0801.

**2nd case:**

If y+ is 50% and y- is 50%, that means both have equal proportion of value,

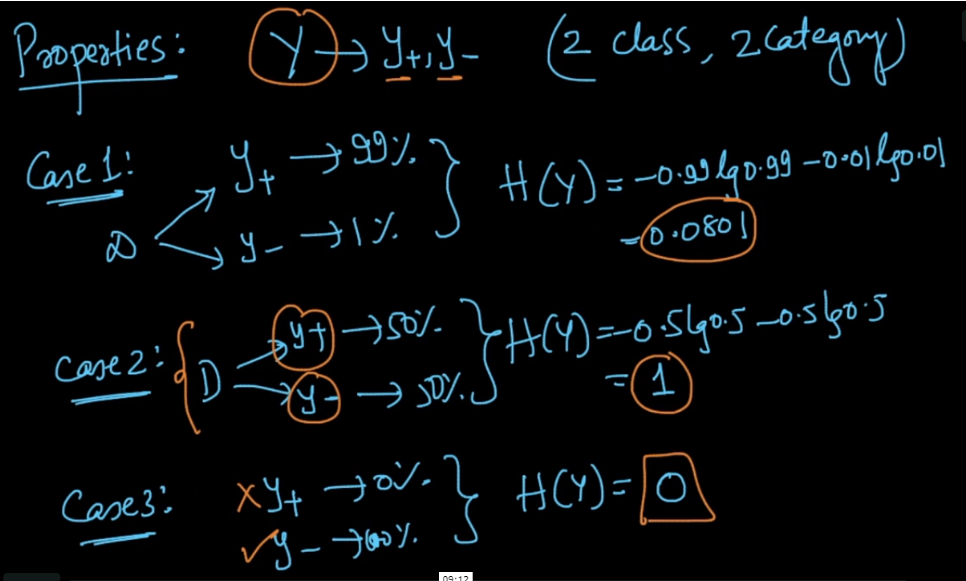
then it’s entropy is 1.

**3rd case:**

If y+ is 0% and y- is 100%, that means there is only false value,

then it’s entropy is 0.

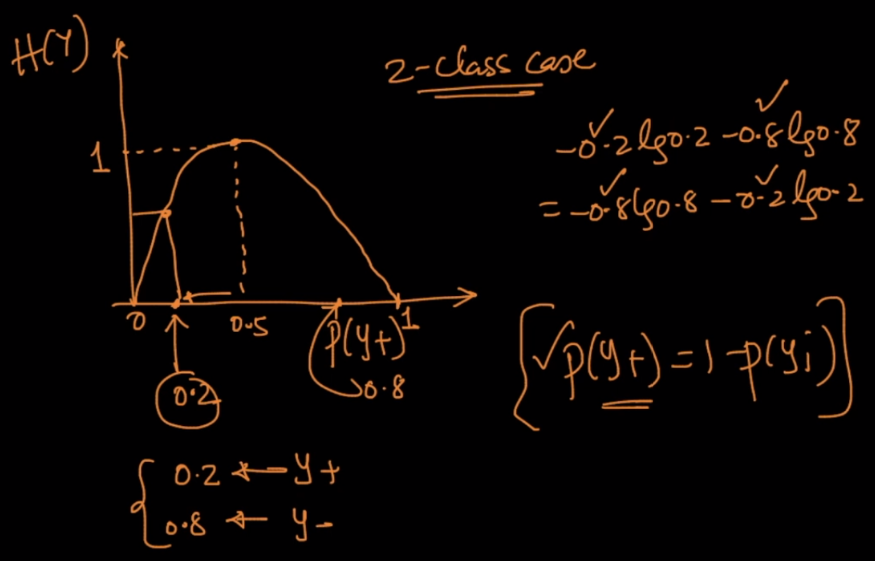
**Conclusion:** The more data in equilibrium(all type of values have equal no. of presence), the more will be entroppy.

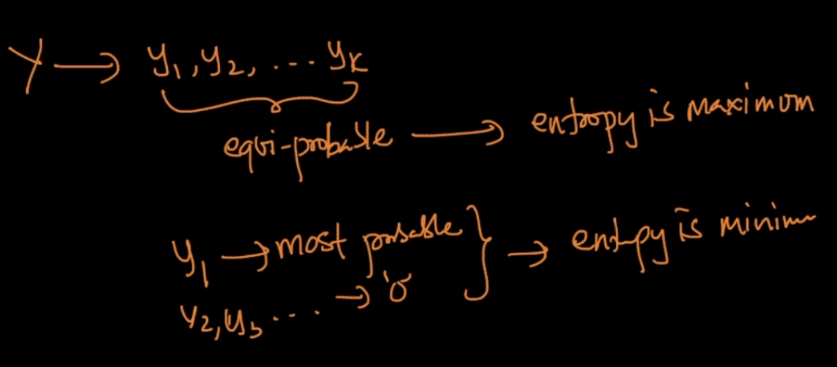


Let’s understand using graph:

Let’s on x-axis we have P(y+) and Y we have entropy, then we can see that when P(y+) = 0.5, this means that P(y-) is having value 0.5, and at this point entorppy is 1 and as we are moving away in any direction entropy is decreasing.

Note: entropy at 0.2 = entropy at 0.8

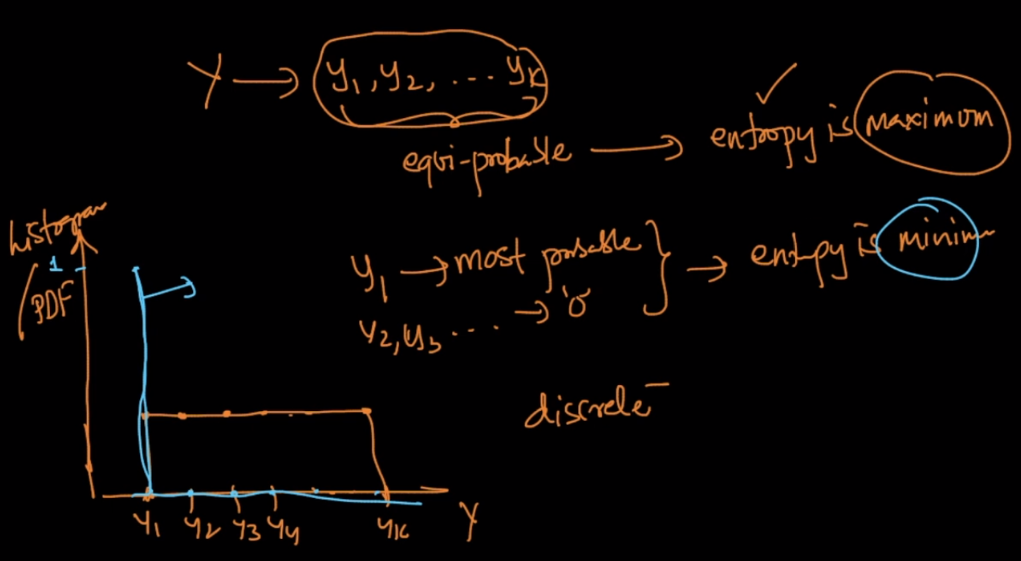




**Entropy graph for discrete variable/categorical feature.**

If one of type of value in that feature have high probability then entroppy value will be low.

If all of type of value in that feature have equal prob then entroppy value will be high or 1.



**Entropy graph for Continuous/numerical feature.**

If a feature has values which is uniformly distributed then entropy will be high or 1.

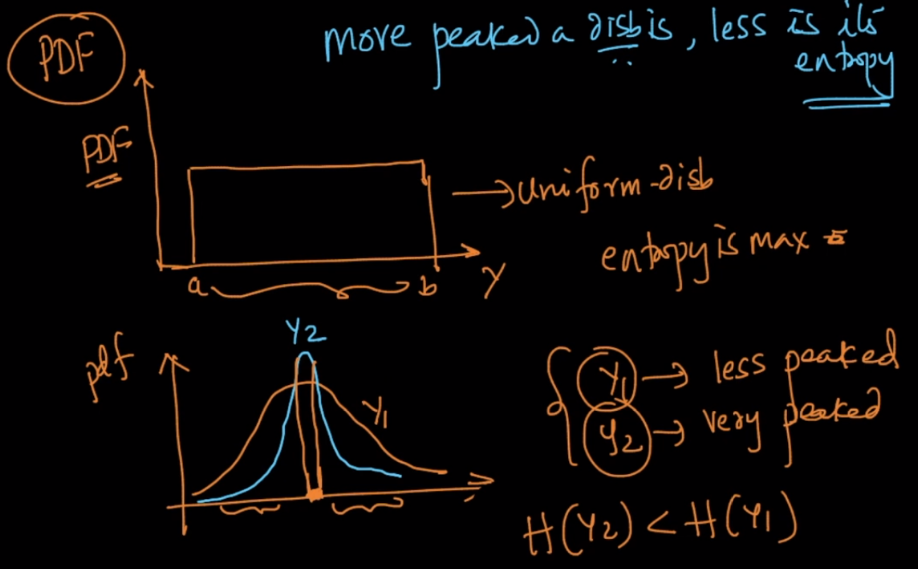
If a feature is not uniformly distributed then:

Entropy(more peaked) < Entropy(less peaked)

Why?

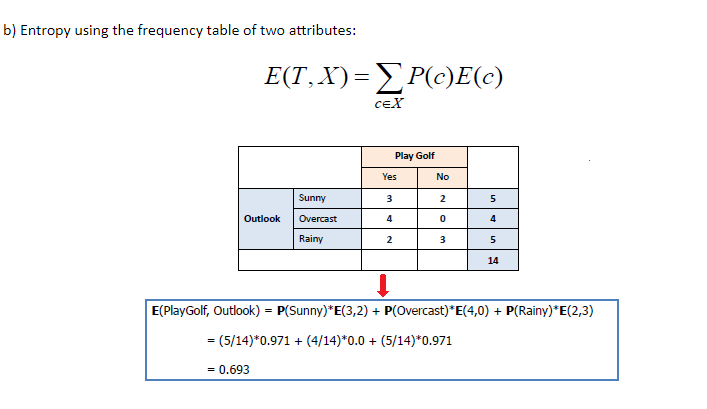
If PDF is more peaked, that means large no. of counts are present in small range value of that feature, and for all other values, there are very less count and hence entropy will be less

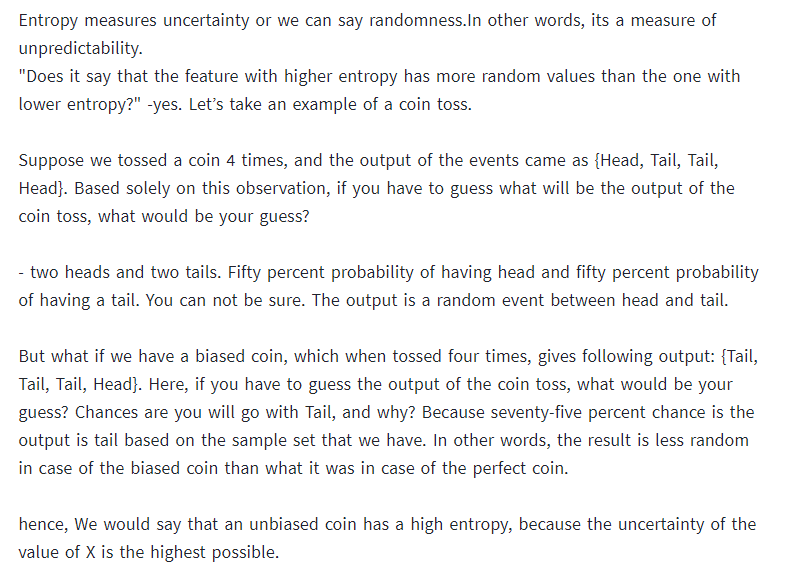
If PDF is less peaked, the counts of values are spread over large range of value of that feature, that means it’s tending toward equilibrium or uniform distribution and hence entropy will be greater.

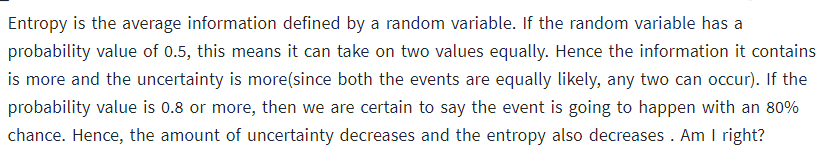


Till now, we calculate entropy for the output variable or feature, but what about independent feature, we’ll do following thing for that: we’ll multiply prob of each type of value for that feature with entropy of feature calculated by separating feature on basis of type of value, and sum that up to find the entropy for that feature. We’ll look this in next chapter

Refer this to know more: <https://www.saedsayad.com/decision_tree.htm>







**Diff b/w variance and entropy**

