Probabilities and Statistics

hbp253

January 2020

1 Likelihood vs Probability Density Functions

Probability density is how likely a set of points are likely to occur given parameter x
Likelihood takes a data set and represents the likeliness of different parameters for your distribution

2 Cohen's d

You can compare the difference between two groups to the variability within groups with Cohen's d:

$$d = \frac{\overline{x_1} - \overline{x_2}}{S}$$

S in this case would be a "pooled standard deviation". Which is a weighed variance between the means:

$$S = \sqrt{\frac{n_1 * var_1 + n_2 * var_2}{n_1 + n_2}}$$

3 PMF

Probability Mass Function or PMF, represent the probability for each unique value. It is normalized by n (the size of the set)

3.1 The Class Size Paradox

Let there be x number of classes for each range of class sizes,
If each one of the student in the class is asked to count the class size,
Then the PMF is multiplied by x again before it is normalized.
Calculating the mean given PMF:

$$\overline{x} = \sum_i p_i x_i,$$

4 CDF

With PMFs, it is difficult to compare two distributions and find meaningful differences visually.
CDFs or Cumulative Distribution Functions are able to show the differences in the distribution,
as it shows the