A secondary way which a distribution can be represented is a probability mass function (PMF).

Probability - is a frequency, a rate of occurrence, relative to a sample size - a total experiment

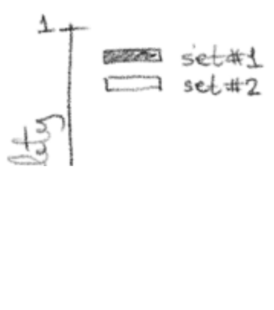
Remember a frequency is the $\boxed{\text{number}}$ of times a value appears in a sample.

$$\text{Freq} \xrightarrow[\text{by sample size}]{\text{divide}} \text{Probabilities}$$

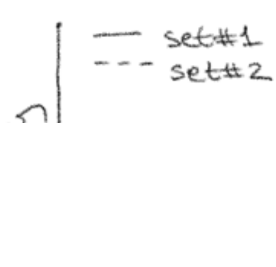
Maintaining a common basis is known as normalization. Usually probability sums to (1).

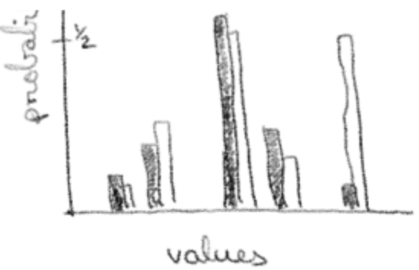
The advantage of a PMF graph is that it obscures differences in sample size - presenting the dispersion of outcomes. This can obfuscate the nature of the data.

PMF Bar Graph



PMF Step Function





"better for small number of values/ outcomes"

"better for many values in an experiment"

Histographs and PMFs are useful ways of representing distributions for exploratory purposes; with goal of identifying patterns or relationships.

Once a pattern or relationship has tentatively been identified other vizualization can be used to better express the theme.

Differences in histographs or PMFs may not be statisticall significant.

A PMF can be afflicted by bias. To counteract that bias, one can take the scaled-by-bias then-normalize PMF and express a new PMF with probabilities inversely proportional to the bias with normalization.

Using this technique a distribution can be gleaned from random samples, removing bias, then accounting for a margin of error.

Mean, variance & std. deviation of PMFs

$$\text{mean} = \bar{x}$$
$$= \sum_i p_i x_i \,,$$

where p_i - is the probability
x_i - is the value

$$\text{sample variance} = S^2$$
$$= \sum p_i (x_i - \bar{x})^2$$

$$\text{sample deviation} = S$$
$$=$$