# ColumbiaX: Machine Learning: Week1 : A probabilistic model

To distinguish whether a model is a probabilistic model or not, one only needs to know the presence or absence of a probabilistic distribution in this model. A probabilistic model is simply a set of distributions on our dataset.
When using a probabilistic model, we usually make the IID assumption, meaning all datapoints in the dataset are generated independently without referring to what values of any other datapoints are and identically from the same distribution, using the same parameter $\theta$ for this specific form distribution. Note that there are many forms of distributions, e.g., binomial, gaussian. But a distribution family refers to one specific form of distribution with $\theta$ unknown.Usually denoted as $\mathbf{p}(\cdot)$. This assumption allows us to write the joint distribution of n datapoints in this way :

$$\mathbf{p}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n|\theta) = \prod_{i=1}^{n} \mathbf{p}(\mathbf{x}_i|\theta)$$