Summary of "A Tutorial for Information Theory in Neuroscience"

1. Why it is beneficial in neuroscience
    a. Multivariate nature of neuroscientific experiments
    b. There are many dependencies between variables and complex data in neuroscience, and information theory can be applied to any mixture of data type
    c. Information theory is well-suited for analyzing and tackling non-linear interactions and relationships (as well as linear)
    d. Model independence – not required to hypothesize a specific structure to the interactions between variables in a data set to use information theory
    e. In a place where creating precise models is extremely difficult, information theory allows for restricting the space of possible models
2. How we can use it
    a. To use information theory, we must first have probability functions/distributions
        i. **Conditional probability distribution**: describes the likelihood to obtain outcomes of certain variables assuming the other variable is known
        ii. **Joint probability distribution:** describes a probability distribution that exists when 2 or more variables exist in a system
        iii. **Normal probability distribution:** probability distribution of a single variable
    b. To get to a place where we can apply IT, we must discretize our data (some experimental data is continuous)
        i. We can do this by binning
            1. Binning is the process of splitting up continuous data into discretized "chunks"
                a. Chunking can be done on the bases of width or count depending on the goal of the experimentation (should not be done in a way that maximizes significance)
3. Information Theory and Entropy
    a. Entropy (or Shannon Entropy) measures the uncertainty contained in a variable and is the fundamental unit of information theory
    b. Characterized by summations of probabilities and base-2 logarithms to form output units of "bits"
    c. Joint entropy is the entropy of a system containing >2 variables
    d. Conditional entropy is an entropic value based on the dependency of one variable to another, or of one variable to many