A taste of information theory and mutation in cancer

Ethan Ashby
Pomona College

September 18, 2020

Mutual Information measures the mutual dependence between two random variables

Pomona College

Given two random variables X and Y, **Mutual information** measures the "amount of information" that X and Y share.

Definition

Given two random variables X and Y, the mutual information, $I(X;Y)$, is defined as:

$$I(X;Y) = H(X) - H(X|Y)$$

Where H represents the entropy (uncertainty) of a random variable.

Intuitively, this definition yields the **reduction in uncertainty** in X when we observe Y.

Cancer is a mutational malady

- Cancer occurs when mutations in DNA cause deregulation of cell cycle control which leads to unchecked division and growth.
- Mutation in cancer is **extremely** rich and heterogeneous.[1]
- We can leverage the heterogeneity in mutation for important clinical tasks: e.g. classifying a cancer's **the tissue of origin**.

[1] Michael S Lawrence et al. "Mutational heterogeneity in cancer and the search for new cancer-associated genes". In: *Nature* 499 (2013), pp. 214–218. DOI: 10.1038/nature12213.

We can use mutual information to tissue type discriminating mutation patterns

Pomona College

Using some fun statistical methods, we can calculate probabilities of encountering mutations in particular regions of the genome under different tissue types:

$$KRAS = \begin{matrix} A & B & C & ... & N \\ (0.1 & 0.01 & 0.01 & ... & 0.9) \end{matrix}$$

Mutual information quantifies the tissue-specificity of these probabilities. The mutual information can tell us whether a mutation in this region is a good predictor of tissue type.