Contrast Information in IDyOMS

- To play nicely with the current estimators in IDyOMS, we assume the following.
  - The past X corresponds to all the past symbols.
    - Not just the previous symbol.
  - The present Y corresponds to the current symbol.
  - The future Z corresponds to the next symbol.
    - Not all the future symbols.
- In order to calculate the different variants of contrast information we have to rewrite them to work with the estimator in IDyOMS.
  - These three distributions are easy to evaluate as is:
    - $p(z|x,y)$, $p(y|x)$, $p(z|y)$
  - These two distributions can be calculated using the previous three:
    - $p(z|x)$, $p(y|x,z)$

$$p(z|x) = \sum_{y \in \mathcal{Y}} p(y|x)p(z|x,y)$$

$$p(y|x,z) = \frac{p(y|x)p(z|x,y)}{p(z|x)}$$

- With these five distributions, we can calculate all three forward contrast variants using the existing estimator in IDyOMS.
- A few gotchas
  - The underlying model must be the same for all of these distributions.
    - More concretely, they must all use the same tally.
  - For instance, the estimate $p(y|x)$ is not simply the previous estimate.
    - It must use the same model (e.g. tally) as $p(z|x,y)$

Forward Predictive Contrast

$$I[Z:y|x] = \sum_{z \in \mathcal{Z}} p(z|x,y) \log \frac{p(z|x,y)}{p(z|x)}$$

- Straightforward estimate of next symbol: $p(z|x,y)$
- Straightforward estimate of current symbol: $p(y|x)$
- Estimate without the present: $p(z|x)$
  - Equivalent to the weighted average over all possible current symbols of the estimate of the next symbol.

$$p(z|x) = \sum_{y \in \mathcal{Y}} p(y|x)p(z|x,y)$$

Forward Connective Contrast

$$I[Z : x|y] = \sum_{z \in \mathcal{Z}} p(z|x, y) \log \frac{p(z|x, y)}{p(z|y)}$$

- Straightforward estimate of next symbol: $p(z|x, y)$
- Estimate with only the present: $p(z|y)$

Forward Reflective Contrast

$$I[Y : z|x] = \sum_{y \in \mathcal{Y}} p(y|x, z) \log \frac{p(y|x, z)}{p(y|x)}$$

- Straightforward estimate of current symbol: $p(y|x)$
- Estimate of current symbol given the next symbol and previous symbols: $p(y|x, z)$
    - Equivalent to the following in forms that are easier to use.

$$p(y|x, z) = \frac{p(y|x)p(z|x, y)}{p(z|x)}$$

- Forward Reflective Contrast can be rewritten as follows.
    - Looks similar to forward predictive contrast.
    - Except that the expectation is different.

$$I[Y : z|x] = \sum_{y \in \mathcal{Y}} p(y|x, z) \log \frac{p(z|x, y)}{p(z|x)}$$

Backward Variants

- Backward Reflective Contrast

$$I[Y : x|z] = \sum_{y \in \mathcal{Y}} p(y|x, z) \log \frac{p(y|x, z)}{p(y|z)}$$

- Backward Connective Contrast

$$I[X : z|y] = \sum_{x \in \mathcal{X}} p(x|y, z) \log \frac{p(x|y, z)}{p(x|y)}$$

- Backward Predictive Contrast

$$I[X : y|z] = \sum_{x \in \mathcal{X}} p(x|y, z) \log \frac{p(x|y, z)}{p(x|z)}$$

- These required the following backward distributions:
    - $p(x|y, z), p(x|y), p(x|z), p(y|z)$

    - Which can be converted to forward distributions and unconditional distributions with the following.

$$p(x|y,z) = \frac{p(z|x,y)p(y|x)p(x)}{p(z|y)p(y)}$$

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$$p(x|z) = \frac{p(z|x)p(x)}{p(z)}$$

$$p(y|z) = \frac{p(z|y)p(y)}{p(z)}$$

- Potential issues
    - The unconditional distributions for $p(y)$ and $p(z)$ should be easy.
        - They are just one-grams.
    - The unconditional distribution for $p(x)$ can be evaluated with the chain rule.
    - The main issue is evaluating the expectation over $\mathcal{X}$.
        - Since the size of the domain grows exponentially with the context length.