Updated Version of the Jern & Kemp (2013) Model

Working backwards, what is needed for simulation of exemplar generation is the probability of generating a stimulus x given exposure to members of the target category x_b:

$$p(x|x_b) = ? \tag{1}$$

where x_b may be empty. Jern & Kemp's model achieves this using a generative process. Category members (x_a or x_b; more generally written as x_y) are assumed to have been generated using an underlying category distribution (specifically, multivariate normal):

$$x_y \sim Normal(\mu_y, \Sigma_y) \tag{2}$$

$p(x|x_y)$ is proportional to the candidate's density under the target category's distribution. Thus, obtaining the category distribution parameters (μ_y, Σ_y) is key for generation.

In Jern & Kemp's original model, category distribution parameters are inferred exclusively from exposure to the category's members. To better simulate our paradigm, category distribution parameters (specifically, the Σ_y parameter) in our implementation are also affected by a higher level domain distribution (specifically, Inverse-Wishart), from which Σ_y are assumed to be generated:

$$\Sigma_y \sim Wishart^{-1}(\Sigma_D, df_D) \tag{3}$$

This document describes how we have obtained the relevant distributions.

Computing μ_y

Because Jern & Kemp assume that the location of x_y is independent of contrast categories, the category μ_y can be inferred exclusively based on exposure to the category's members. Assuming (μ_y, Σ_y) are Normal-Inverse-Wishart distributed (unknown mean, unknown variance):

$$\mu_y = \frac{\kappa\mu_0 + n_y\bar{x_y}}{\kappa + n_y} \tag{4}$$

where:

- μ_0 is the prior mean along p dimensions. Here we set it to the middle of the space.
- κ is a scalar hyper-parameter, roughly weighting the importance of μ_0. κ must be greater than zero.
- n_y is the number of observations in x_y
- $\bar{x_y}$ is the sample mean along p dimensions

In the case of a populated class, μ_y ends up lying somewhere between μ_0 and $\bar{x_y}$, depending on κ_0 and n_y. In the case of an empty class, $n_y = 0$, thus Equation 4 reduces to $\mu_y = \mu_0$. Because we set μ_0 to the center of the space, this outcome is the same as if we had integrated over all possible μ_y.

Computing Σ_D

Unlike μ_y, Σ_y cannot be computed considering only the members of category y. Instead, Σ_y is influenced both by the distribution of x_y and by members of other categories through Σ_D.

Σ_D is inferred based on the observed (empirical) category covariances C_y. We assume these covariances to be Wishart-distributed, and so Σ_D can be computed as:

$$\Sigma_D = \Sigma_0 + \sum_y C_y \tag{5}$$

Σ_0 is a p-by-p prior covariance matrix. We use a p-dimensional identity matrix I_p multiplied element-wise against a free parameter, γ, controlling the amount of variance assumed by the prior:

$$\Sigma_0 = I_p \gamma \tag{6}$$

Note

In actuality, we ended up doing a bunch of matrix inverting that I do not understand:

```
Domain_Sigma = np.linalg.inv(Prior_Sigma)
for y in ncategories:
  Domain_Sigma += np.linalg.inv(Category_Sigmas[:,:,y])
Domain_Sigma = np.linalg.inv(Domain_Sigma)
```

This is definitely not in the math, and I think it has something to do with Wishart vs. Inverse-Wishart. I think it also had something to do with the issue that its weird to just take the sum of the covariance matrices: the domain ends up assuming much larger covariances compared to any of the categories. We've never nailed it down. **Joe**, do you know what this is about?

Computing Σ_y

In previous versions, we sampled $\Sigma_y \sim Wishart^{-1}(\Sigma_D, df_D)$, with which we sampled $x \sim Normal(\mu_y, \Sigma_y)$. That was the approach taken by Jern & Kemp. Sampling x, however, is not the primary goal of the model – the main purpose of the model is to predict $p(x|x_b)$, and from those data we can sample x.

We could monte-carlo sample x as described above to obtain $p(x|x_b)$. But this is computationally intensive, as each trial would need a separate sampler. Instead, it would be better to develop a method to infer Σ_y without sampling.

Assuming (μ_y, Σ_y) are Normal-Inverse-Wishart distributed, Σ_y can be computed as:

$$\Sigma_y = [\Sigma_D \nu + C_y + \frac{\kappa n_y}{\kappa + n_y}(\bar{x_y} - \mu_y)(\bar{x_y} - \mu_y)^T](\nu + n_y)^{-1} \tag{7}$$

κ, $\bar{x_y}$, C_y, n_y, μ_0, are the same values as described above. ν is an additional free parameter, weighting the importance of Σ_D. ν must be greater than $p-1$. When x_b is empty, Equation 7 reduces to $\Sigma_y = \Sigma_D$.

Computing response probabilities $p(x|x_b)$

x are assumed to be drawn from a the distribution given by $Normal(\mu_y, \Sigma_y)$. Thus,

$$p(x) \propto Normal(x|\mu_b, \Sigma_b) \tag{8}$$

In practice, $p(x)$ is computed by first obtaining the relative density of every possible generation candidate x_i under the category distribution. The end probability is a normalization of these values:

$$p(x) = \frac{exp(\theta Normal(x|\mu_b, \Sigma_b))}{\sum_i exp(\theta Normal(x_i|\mu_b, \Sigma_b))} \tag{9}$$

where θ is a response determinism parameter.

Description of free parameters

- κ. Scalar, $\kappa > 0$. Weights the importance of μ_0 in inferring category μ_y.
- γ. Scalar, $\gamma > 0$. Weights the importance of Σ_0 in inferring the domain Σ_D.
- ν. Scalar, $\nu > p - 1$. Weights the importance of Σ_D in inferring the domain Σ_y.
- θ. Scalar, $\theta > 0$. Response determinism parameter.