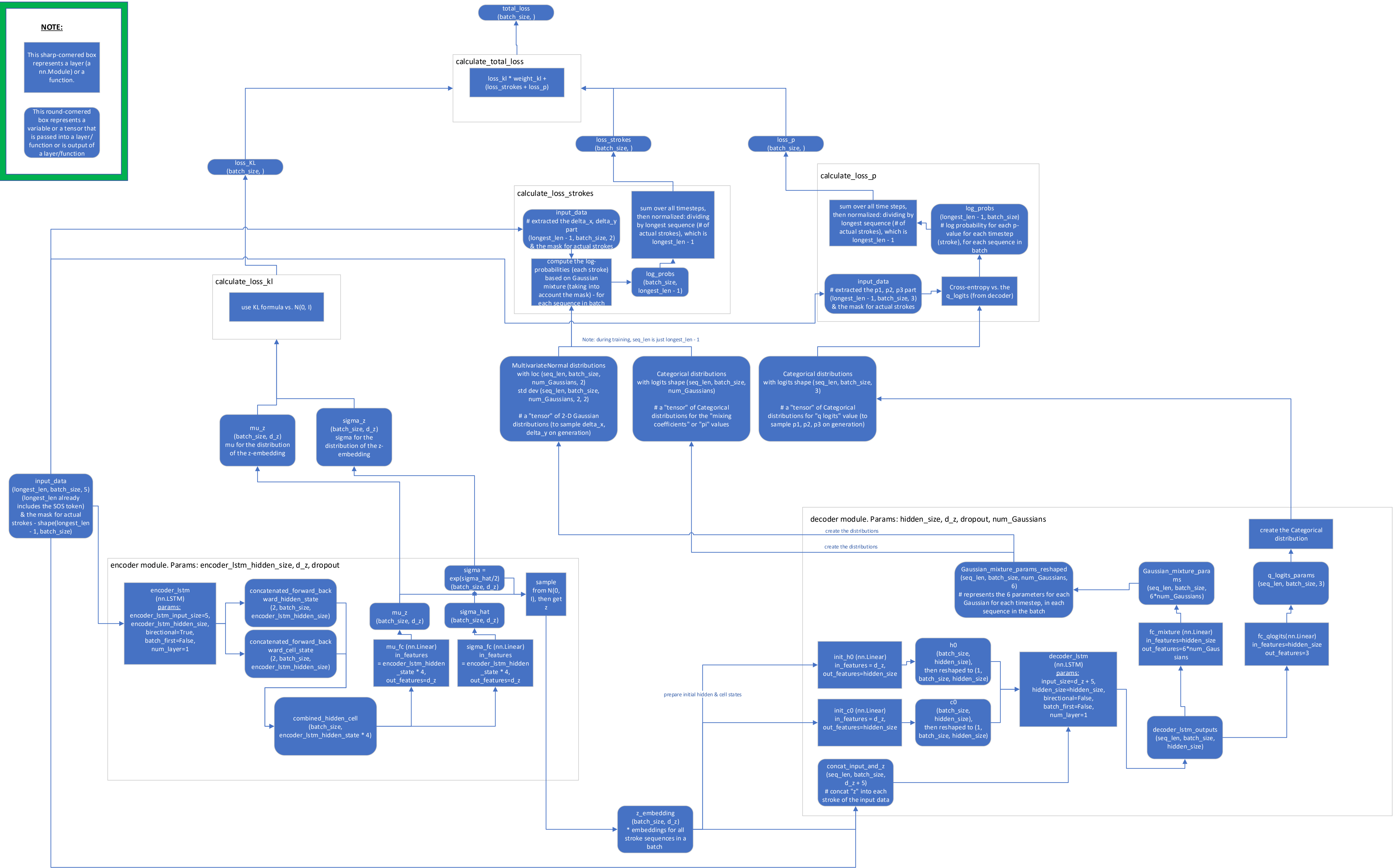


**NOTE:**

This sharp-cornered box represents a layer (a `nn.Module`) or a function.

This round-cornered box represents a variable or a tensor that is passed into a layer/function or is output of a layer/function



During training: leave out the last stroke. So input into decoder has length "seq\_len" which equals longest\_len - 1. Note that during generation, seq\_len = 1.