The K-class categorical distribution has probability mass function

$$f\left(y|p\right) = \prod_{k \in \{1,\ldots,K\}} p_k^{y_k}$$

where $y \in \{0,1\}^K : \sum_{k=1}^K y_k = 1$ is the one-hot encoding of the observed class.
The $(K-1)$-simplex is $\Delta^{K-1} = \{p \in \mathbb{R}^K : p_k \geq 0 \; \forall k, \; \sum_{k=1}^K p_k = 1\}$
Given a model $m : \mathbb{X} \to \Delta^{K-1}$, the log loss on a set of observations $\{x_i, y_i\}_{i=1}^N$ is

$$LL = -\frac{1}{N} \sum_{i=1}^{N} \ln \left(\prod_{k=1}^{K} \left[m_k \left(x_i \right) \right]^{y_k} \right) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_k \ln \left(m_k \left(x_i \right) \right)$$