

The log-loss framework

- ▶ Algorithm A predicts a sequence c^1, c^2, \dots, c^T over alphabet $\Sigma = \{1, 2, \dots, k\}$
- ▶ The prediction for the c^t th is a distribution over Σ :
 $\mathbf{p}_A^t = \langle p_A^t(1), p_A^t(2), \dots, p_A^t(k) \rangle$
- ▶ When c^t is revealed, the loss we suffer is $-\log p_A^t(c^t)$
- ▶ The **cumulative log loss**, which we wish to minimize, is
 $L_A^T = -\sum_{t=1}^T \log p_A^t(c^t)$
- ▶ $\lceil L_A^T \rceil$ is the code length if A is combined with arithmetic coding.