

## log loss encourages unbiased prediction

- ▶ Suppose the source is random and the probability of the next outcome is  $p(c_t | c_1, c_2, \dots, c_{t-1})$
- ▶ Then the prediction that minimizes the log loss is  $p(c_t | c_1, c_2, \dots, c_{t-1})$ .
- ▶ Note that when minimizing expected number of mistakes, the best prediction in this situation is to put all of the probability on the most likely outcome.
- ▶ There are other losses with this property, for example, square loss.