

Arithmetic Coding (basic idea)

- ▶ Easier notation: represent characters by numbers
 $1 \leq c_t \leq |\Sigma|$. (English: $|\Sigma| = 26$)
- ▶ message-prefix c_1, c_2, \dots, c_{t-1} represented by line segment $[l_{t-1}, u_{t-1})$
- ▶ Initial segment $[l_0, u_0) = [0, 1)$
- ▶ After observing c_1, c_2, \dots, c_{t-1} , predictor outputs
 $p(c_t = 1 | c_1, c_2, \dots, c_{t-1}), \dots, p(c_t = |\Sigma| | c_1, c_2, \dots, c_{t-1})$,
- ▶ Distribution is used to partition $[l_{t-1}, u_{t-1})$ into $|\Sigma|$ sub-segments.
- ▶ next character c_t determines $[l_t, u_t)$