Generalization of the HMM
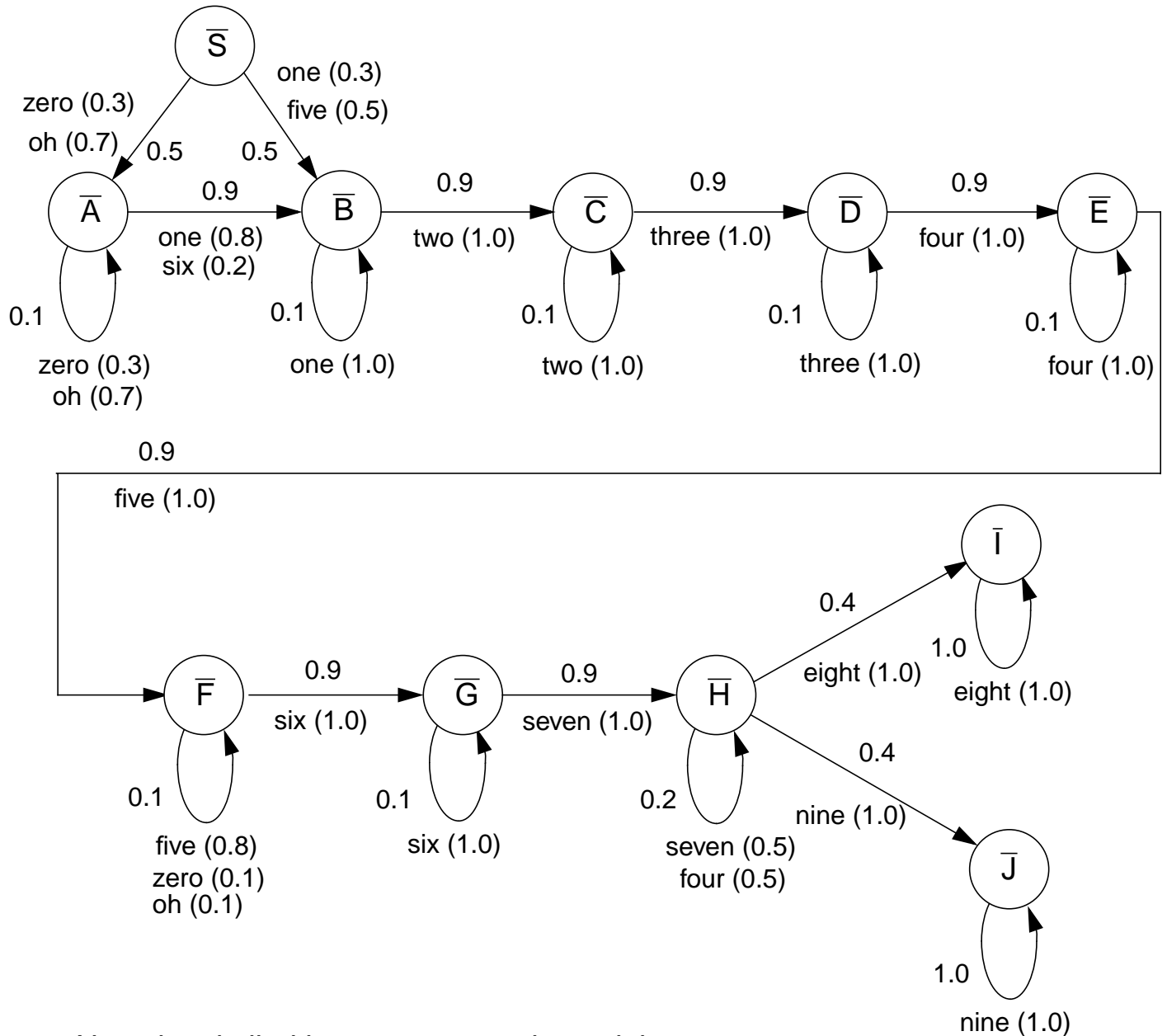
Consider the following state diagram showing a simple language model involving constrained digit sequences:

$\overline{S}$

one (0.3)
five (0.5)

zero (0.3)
oh (0.7)   0.5   0.5

$\overline{A}$ 0.9 $\overline{B}$ 0.9 $\overline{C}$ 0.9 $\overline{D}$ 0.9 $\overline{E}$

one (0.8)   two (1.0)   three (1.0)   four (1.0)
six (0.2)

0.1   0.1   0.1   0.1   0.1

zero (0.3)   one (1.0)   two (1.0)   three (1.0)   four (1.0)
oh (0.7)

0.9

five (1.0)

$\overline{I}$

0.4   1.0
eight (1.0)   eight (1.0)

$\overline{F}$ 0.9 $\overline{G}$ 0.9 $\overline{H}$

six (1.0)   seven (1.0)

0.4

0.1   0.1   0.2   nine (1.0)

five (0.8)   six (1.0)   seven (0.5) $\overline{J}$
zero (0.1)   four (0.5)
oh (0.1)

1.0

nine (1.0)

Note the similarities to our acoustic models.

What is the probability of the sequence "zero one two three four five zero six six seven seven eight eight" ?

How would you find the average length of a digit sequence generated from this language model?

In the terminology associated with formal language theory, this HMM is known as a *finite state automaton*.

The word *stochastic* can also be applied because the transitions and output symbols are governed by probability distributions.

Further, since there are multiple transitions and observations generated at any point in time (hence, ambiguous output), this particular graph is classified as a *nondeterministic* automaton.

In the future, we will refer to this system as a stochastic finite state automaton (FSA or SFSA) when it is used to more *linguistic* information.

We can also express this system as a *regular grammar*:

$\overline{S} \xrightarrow{p_1}$ zero, $\overleftarrow{A}$　　$\overline{D} \xrightarrow{p_{13}}$ three, $\overleftarrow{D}$　　$\overline{H} \xrightarrow{p_{25}}$ eight, $\overleftarrow{I}$

$\overline{S} \xrightarrow{p_2}$ oh, $\overleftarrow{A}$　　$\overline{D} \xrightarrow{p_{14}}$ four, $\overleftarrow{E}$　　$\overline{H} \xrightarrow{p_{26}}$ nine, $\overleftarrow{J}$

$\overline{S} \xrightarrow{p_3}$ one, $\overleftarrow{A}$　　$\overline{E} \xrightarrow{p_{15}}$ four, $\overleftarrow{E}$　　$\overline{I} \xrightarrow{p'_{27}}$ eight, $\overleftarrow{I}$

$\overline{S} \xrightarrow{p_4}$ five, $\overleftarrow{A}$　　$\overline{E} \xrightarrow{p_{16}}$ five, $\overleftarrow{F}$　　$\overline{I} \xrightarrow{p''_{27}}$ eight.

$\overline{A} \xrightarrow{p_5}$ zero, $\overleftarrow{A}$　　$\overline{F} \xrightarrow{p_{17}}$ zero, $\overleftarrow{F}$　　$\overline{I} \xrightarrow{p'_{28}}$ nine, $\overleftarrow{J}$

$\overline{A} \xrightarrow{p_6}$ oh, $\overleftarrow{A}$　　$\overline{F} \xrightarrow{p_{18}}$ oh, $\overleftarrow{F}$　　$\overline{I} \xrightarrow{p''_{28}}$ nine.

$\overline{A} \xrightarrow{p_7}$ one, $\overleftarrow{B}$　　$\overline{F} \xrightarrow{p_{19}}$ five, $\overleftarrow{F}$

$\overline{A} \xrightarrow{p_8}$ six, $\overleftarrow{B}$　　$\overline{F} \xrightarrow{p_{20}}$ six, $\overleftarrow{G}$

$\overline{B} \xrightarrow{p_9}$ one, $\overleftarrow{B}$　　$\overline{G} \xrightarrow{p_{21}}$ six, $\overleftarrow{G}$

$\overline{B} \xrightarrow{p_{10}}$ two, $\overleftarrow{C}$　　$\overline{G} \xrightarrow{p_{22}}$ seven, $\overleftarrow{H}$

$\overline{C} \xrightarrow{p_{11}}$ two, $\overleftarrow{C}$　　$\overline{H} \xrightarrow{p_{23}}$ seven, $\overleftarrow{H}$

$\overline{C} \xrightarrow{p_{12}}$ three, $\overleftarrow{D}$　　$\overline{H} \xrightarrow{p_{24}}$ four, $\overleftarrow{H}$

Note that rule probabilities are not quite the same as transition probabilities, since they need to combine transition probabilities and output probabilities. For example, consider p_7:

$$p_7 = (0.9)(0.8)$$

In general,

$$P(\underline{y} = y_k | \underline{x} = x_i) = \sum_j a_{ij} b(k)$$

Note that we must adjust probabilities at the terminal systems when the grammar is nondeterministic:

$$p_k = p'_k + p''_k$$

to allow generation of a final terminal.

Hence, our transition from HMMs to stochastic formal languages is clear and well-understood.

What types of language models are used?

- No Grammar (Digits)
- Sentence pattern grammars (Resource Management)
- Word Pair/Bigram (RM, Wall Street Journal)
- Word Class (WSJ, etc.)
- Trigram (WSJ, etc.)
- Back-Off Models (Merging Bigrams and Trigrams)
- Long Range N-Grams and Co-Occurrences (SWITCHBOARD)
- Triggers and Cache Models (WSJ)
- Link Grammars (SWITCHBOARD)

How do we deal with OOV and dysfluencies?