Number:

Problem	Points	Score
1 (a)	10	
1 (b)	15	
1 (c)	15	
1 (d)	15	
1 (e)	15	
2 (a)	20	
2 (b)	10	
Total	100	

Notes:

1. The exam is closed books and notes. You are allowed one 8 1/2” x 11” double-sided sheet of notes.

2. Please indicate clearly your answer to the problem by some form of highlighting (underlining).

3. Your solutions must be legible and easy to follow. If I can’t read it or understand it, it is wrong. Random scribbling will not receive credit.

4. Please show ALL work. Answers with no supporting explanations or work will be given no credit.

5. Several problems on this exam are fairly open-ended. Since the evaluation of your answers is obviously a subjective process, we will use a market place strategy in determining the grade. Papers will be rank-ordered in terms of the quality of the solutions, and grades distributed accordingly.

1. Before his death, Chairman Mao devised a plot to overthrow the western world by designing a coin called the "Jonnie," shown to the right. This coin had three sides: "Top" denoted "T", "Bottom" denoted "B," and "Sides", denoted "S". The plot involved playing the ancient Chinese game known as JonGo, and betting on the outcome. Chairman Mao demonstrated this game to his top military advisors by flipping the coin 10 consecutive times, and doing three separate trials. He generated the following observations:

Trial No. 1: BBBTTBBBTT
Trial No. 2: TTBBBTTBBB
Trial No. 3: BBBBBBTTTS

Your job is to model this system using a speech recognition system. Let's begin with some simple calculations.

(a) Estimate a unigram probability density function for the observable symbols in this system (the state of the coin after each toss). Then compute the entropy: $H(x) = -\sum p(x)\log_2 p(x)$.

(b) Estimate a bigram probability density function for this coin. (**Do not take into account an end of string symbol**.) Compute the entropy. Compare and contrast this number with the answer to part (a). Why are they different?

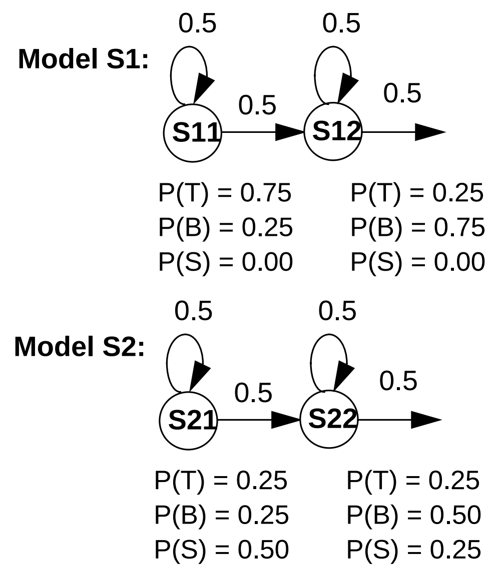
(c) Smooth the bigram probability distribution using any two smoothing techniques discussed in class. Compare and contrast the entropies computed for these smoothed distributions. Why are they different?

(d) Describe a language model that will better model the data for this experiment. Justify your answer using as many detailed calculations as possible.

(e) Compute the training set perplexity using the unigram language model of (a). Compare and contrast this value with the average branching factor. Explain how your observations might be relevant to the speech recognition problem.

2. Some smart American government employees, who work for an agency that doesn't exist, decide they can sabotage this plot by training a speech recognition system on this data. The system they develop looks like this:

Language Model:	Lexicon:	Acoustic Models:
P(T) = 0.50	T -> S1 S2	
P(B) = 0.25	B -> S2 S1	
P(S) = 0.25	S -> S2 S2	



(a) Find the model and state sequences that were most likely to produce the observation sequence "TB".

(b) Estimate the computational complexity of your calculation in part (a) and discuss how you might make it more efficient.