Chain Rule

- P(A,B) =
- P(A,B,C) =
- P(A,B,C,D) =

Chain Rule

- $P(A,B) = P(A)P(B|A) = P(B)P(A|B)$
- $P(A,B,C) = P(A)P(B,C|A) = P(A)P(B|A)P(C|A,B)$
- $P(A,B,C,D) = P(A)P(B|A)P(C|A,B)P(D|A,B,C)$

(Conditional) Independence

- Knowing about A doesn't help you narrow down the likelihood of B taking on a certain value.
    - e.g. A = temperature in Claremont, B = value of the U.S. dollar
    - What does that mean for the chain rule?
- Conditional independence example: Word likelihood and topic

Joint Probabilities

- In the real world, we commonly observe the interaction of many related random variables.

- Language example: Word frequencies and text type; alignments and translation likelihoods

Marginal Probabilities

- For analysis, we commonly want to know about the behavior of a single variable

- “If I pick a random word from a random book at the library, what is the probability that my word is “sauté”?

Putting it Together…

- Is the chance of rain on a given day in Seattle (S) and Claremont (c) independently distributed?

- Hint: Calculate the marginal chance of rain in each city.

s	Rain	Rain	No Rain	No Rain
c	Rain	No Rain	Rain	No Rain
P(S=s,C=c)	0.043	0.387	0.057	0.513

Bayes Rule

- P(A|B) = (P(B|A)P(A))/P(B)

- Why might this be useful to us in MT?

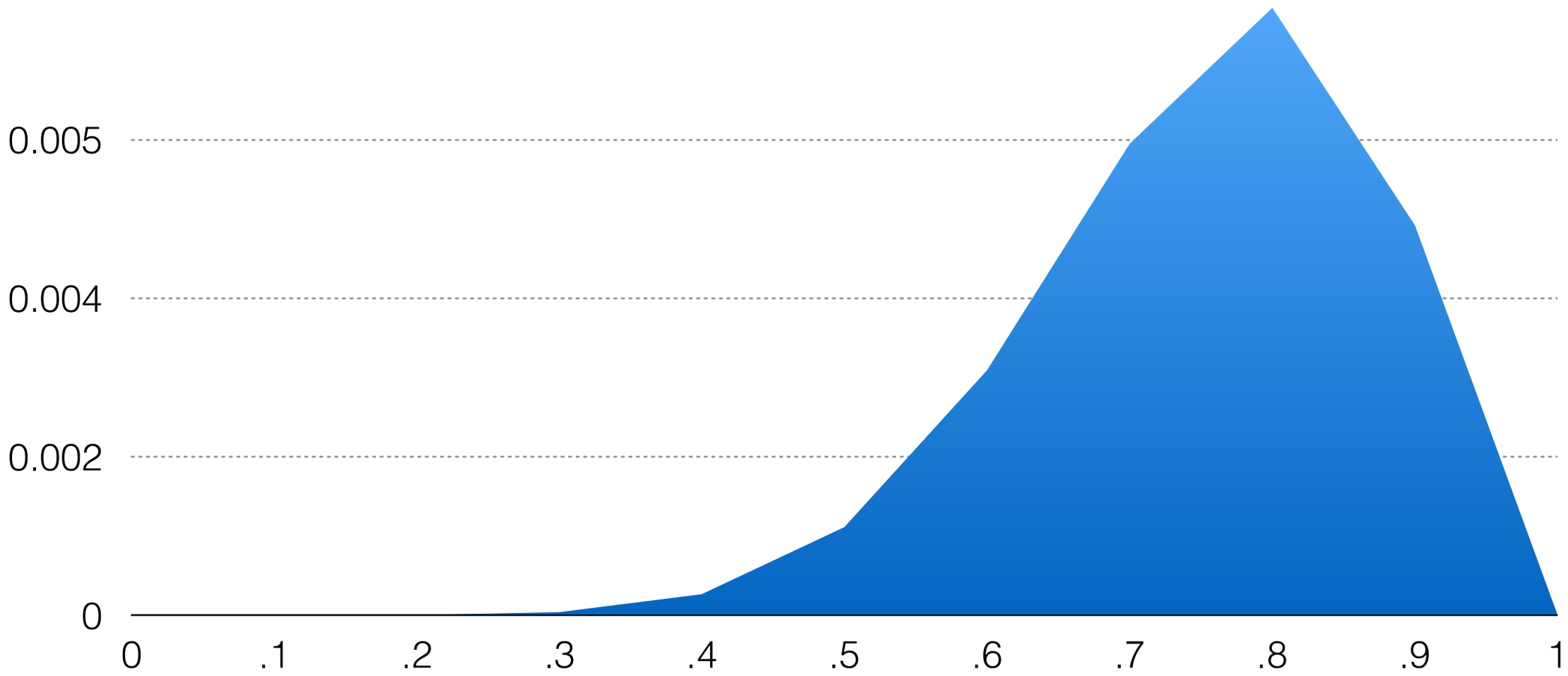
MLE

- Maximum Likelihood Estimate
  - Given observations, estimate parameters
    - Non-language example: Flipping a potentially unfair coin
    - Language example: Given alignments, estimate translation probabilities

MLE Example

- Suppose I toss a coin 10 times
- Result: 8 heads, then 2 tails
    - Sketch likelihood of this result as a function of p(heads)

P(8 heads, 2 tails)



Another MLE Example

- Unknown language alignment
  - Suppose we know the alignments between the words in two sets of words.
  - Then we can estimate the parameters of the model that make that alignment most likely.

uif		qptu		pggjdf		jt		efnpdsbdz	
mb	12007	qptuf	20	pggjdf	39	ftu	5556	eênpdsbujf	144
mf	7812	qptubvy	3	cvsfbv	17	(	769	mb	99
mft	5867	qptuft	2	qptuf	9	t(	428	ef	10
m(	4761	dpvssjfs	2	nboebu	5	bhju	338	eênpdsbujrvf	8

uif qptu pggjdf jt efnpdsbdz

eênpdsbujf mb ftu d(qptuf mb

uif		qptu		pggjdf		jt		efnpdsbdz	
mb	0.4	qptuf	0.7	pggjdf	0.6	ftu	0.8	eênpdsbujf	0.6
mf	0.3	qptubvy	0.1	cvsfbv	0.2	(	0.1	mb	0.4
mft	0.2	qptuft	0.1	qptuf	0.1	t(	0.1	ef	0.03
m(	0.2	dpvssjfs	0.1	nboebu	0.1	bhju	0.05	eênpdsbujrvf	0.03

uif qptu pggjdf jt efnpdsbdz

eênpdsbujf mb ftu d(qptuf mb

Project 1

- Implement word alignment: IBM Model 1 plus one extension
  - A model that prefers to align words close to the diagonal.
  - An HMM alignment model.
  - Combine French-English model and English-French model.
  - Supervised discriminative alignment model
  - Unsupervised discriminative alignment model.