

## Notes 3-18-19

### New Distributions for BRACS Paper

#### New Distributions

I tried out some new distributions to more closely mirror the voter registration context. Here, I am using first name, last name, race, and party. The  $u$  parameters for all of these are somewhat fixed, but the  $m$  can be adjusted to weaken or strengthen the variable. When I tried a 5th variable, I ran into the same problem where there was too much information in the data, so you couldn't clearly see the effects of BRACS.

m and u distributions (New)				
	Cluster 1		Cluster 2	
	m1	m2	u1	u2
fname0	0.05	0.05	0.9900000	0.9900000
fname1	0.95	0.95	0.0100000	0.0100000
lname0	0.05	0.80	0.9900000	0.9900000
lname1	0.95	0.20	0.0100000	0.0100000
race_0	0.02	0.02	0.6666667	0.6666667
race_1	0.98	0.98	0.3333333	0.3333333
party_0	0.02	0.02	0.5000000	0.5000000
party_1	0.98	0.98	0.5000000	0.5000000

## Results

As mentioned before, we see a substantial increase in precision, which also leads to an increase in fmeasure. For the average link probabilities, I split the records by their linkage cluster, looked at the declared matches, and took an average of those posterior match probabilities. On their own, these probabilities do not tell us much, since they could just indicate we are more confident in wrong decisions. However, since we see an increase in fmeasure, I think these numbers are useful. Also, there may be better ways to report or calculate this information, so I'll keep thinking about it.

