Problemset 4, Econ 980w, Spring 2019: Distributional decompositions

1 By hand exercises

Suppose you have the following data-set on education (dropout, high school, college), immigrant status (migrant, native), and wages per hour.

Education	Migrant	Wage
D	M	4
D	M	5
D	N	6
HS	M	5
HS	M	8
HS	N	6
HS	N	7
HS	N	10
C	M	14
C	N	12
C	N	18

1. What is the mean wage of migrants, and of natives? What is the median wage of either group?

2. Suppose migrants were to have the same educational distribution as natives. What would be their mean wage? Their median wage?

2 R exercises

1. Suppose X can take on 4 values, depending on an individual's gender and on whether he/she graduated from college. Suppose you have matrices A^1 and A^0 completely describing $P^1(X, D)$ and $P^0(X, D)$, as well as matrices B^1 and B^0 describing $E^t[Y|X, D]$. What is the dimension of these matrices? Pick some numbers for these matrices A^t and B^t (be careful to make sure that probabilities add up to 1).

2. Write a Matlab script that calculates θ and $E^*[Y]$ based on the matrices A^t and B^t. Where possible, do calculations by matrix multiplication rather than using loops.

3. Next, generate random samples from the distributions $P^t(Y, D, X)$ under the same assumptions, that is with distributions of X and D determined by some matrices A and B that you picked, and for values Y drawn from the distributions

$$Y|X, D \sim N(E^t[Y|X, D], 1).$$

Export a dataset containing draws of the variables t, Y, X, D, and give it to a classmate.

4. Use the data you got from your classmate to estimate $E^*[Y]$, building on the code you wrote for 1. Compare your estimate to the number they got in step one based on the true matrices A and B.