7 Exercises (due Thursday, 29.01.2024, 09:00am)

There is a total of 7 points to achieve on this last exercise sheet. Please upload your solutions in typed or readable (!) handwritten form (scans, photographs) on Moodle and don't forget to flag all the problems you were able to solve. Good luck!

Exercise 7.1 (3P). *Show that*

$$\hat{se}^2 := \sum_{k,l=1}^{n} y_{i_k} y_{i_l} \left(\frac{1}{\pi_{i_k}\pi_{i_l}} - \frac{1}{\pi_{i_k i_l}} \right)$$

is an unbiased estimator for the variance $\mathrm{Var}[\hat{\tau}]$ *of the Horvitz-Thompson estimator* $\hat{\tau} = \sum_{j=1}^{n} \frac{y_{i_j}}{\pi_{i_j}}$ *with* $S = (i_1, \ldots, i_n) \in \mathcal{U}^n$*.*

Exercise 7.2 (4P). *Two-step snowball sampling on a population graph* $G = (V, E)$ *is defined as follows: First, draw* V_0^* *of size* n *uniformly from* V *without replacement. Then add all edges incident to* V_0^* *in order to obtain* E_1^**. Next, add all vertices that are incident to* E_1^* *to obtain* V_1^**. Now, add all edges that are incident to* V_1^* *in order to obtain* E_2^**. Finally, add all vertices that are incident to* E_2^* *to obtain* V_2^**. Set* $V^* = V_2^*$ *and* $E^* = E_2^*$*.*

a) ***(3P)*** *Generate a population graph* $G = (V, E)$ *on* N_V *vertices from the Erdös-Rényi model (that is, each possible edge is included in* E *with a given probability* p *independently of all other edges) with* $p = \frac{\log N_V}{N_V}$*. Conduct a Montecarlo experiment to approximate the vertex inclusion probabilities* π_v*,* $v \in V$*, of two-step snowball sampling. Visualize these inclusion probabilities, ordering the vertices on the* x*-axis according to their degree in the population graph. Try also ordering the vertices according to the size of their second-order neighborhoods in the population graph. For* $v \in V$*, the second-order neighborhood of* v *is defined as* $N_v^{(2)} := \{u \in V : dist(u, v) \leq 2\}$*.*

b) ***(1P)*** *Derive an explicit formula for* π_v*.*

Hint: Don't regenerate the population graph in each Montecarlo iteration. The population is fixed. We are simulating the process of sampling from the population!