

Unequal probability sampling designs

January 12, 2021

1) Some examples of using maximum entropy sampling design and related functions:

a) First example

Sample of Belgian municipalities, sample size 50

```
> data(belgianmunicipalities)
> attach(belgianmunicipalities)
> n=50
```

Inclusion probabilities proportional to the 'averageincome' variable

```
> pik=inclusionprobabilities(averageincome,n)
```

Draw a sample

```
> s=UPmaxentropy(pik)
```

The sample is

```
> as.character(Commune[s==1])
```

Joint inclusion probabilities

```
> pi2=UPmaxentropypi2(pik)
```

Check the result

```
> rowSums(pi2)/pik/n
> detach(belgianmunicipalities)
```

b) Second example

Selection of samples from Belgian municipalities data set, sample size 50. Once matrix q is computed, a sample is quickly selected. Simulations can be run to compare the results.

```

> data(belgianmunicipalities)
> attach(belgianmunicipalities)
> pik=inclusionprobabilities(averageincome,50)
> pik=pik[pik!=1]
> n=sum(pik)
> pikt=UPMEpiktildefrompik(pik)
> w=pikt/(1-pikt)
> q=UPMEqfromw(w,n)

```

Draw a sample using the q matrix

```

> UPMEsfromq(q)
>

```

Simulations to check the sample selection; the difference between pik and the computed inclusion prob. tt is almost 0.

```

> sim=10000
> N=length(pik)
> tt=rep(0,N)
> for(i in 1:sim) tt = tt+UPMEsfromq(q)
> tt=tt/sim
> max(abs(tt-pik))
> detach(belgianmunicipalities)

```

2) This is an example of unequal probability (UP) sampling functions: selection of samples using the Belgian municipalities data set, with equal or unequal probabilities, and study of the Horvitz-Thompson estimator accuracy using boxplots. The following sampling schemes are used: Poisson, random systematic, random pivotal, Tillé, Midzuno, systematic, pivotal, and simple random sampling without replacement. Monte Carlo simulations are used to study the accuracy of the Horvitz-Thompson estimator of a population total. The aim of this example is to demonstrate the effect of the auxiliary information incorporation in the sampling design. We use:

- some π ps sampling designs with Horvitz-Thompson estimation, using in the sampling design the information on size measures of population units;
- simple random sampling without replacement with Horvitz-Thompson estimation, where no auxiliary information is used.

```

> b=data(belgianmunicipalities)
> pik=inclusionprobabilities(belgianmunicipalities$Tot04,200)
> N=length(pik)
> n=sum(pik)

```

Number of simulations (for an accurate result, increase this value to 10000):

```
> sim=10
> ss=array(0,c(sim,8))
```

Defines the variable of interest:

```
> y=belgianmunicipalities$TaxableIncome
```

Simulation and computation of the Horvitz-Thompson estimator:

```
> ht=numeric(8)
> for(i in 1:sim)
+ {
+   cat("Step ",i,"\n")
+   s=UPpoisson(pik)
+   ht[1]=HTestimator(y[s==1],pik[s==1])
+   s=UPrandomsystematic(pik)
+   ht[2]=HTestimator(y[s==1],pik[s==1])
+   s=UPrandompivotal(pik)
+   ht[3]=HTestimator(y[s==1],pik[s==1])
+   s=UPtille(pik)
+   ht[4]=HTestimator(y[s==1],pik[s==1])
+   s=UPmidzuno(pik)
+   ht[5]=HTestimator(y[s==1],pik[s==1])
+   s=UPsystematic(pik)
+   ht[6]=HTestimator(y[s==1],pik[s==1])
+   s=UPpivotal(pik)
+   ht[7]=HTestimator(y[s==1],pik[s==1])
+   s=srswor(n,N)
+   ht[8]=HTestimator(y[s==1],rep(n/N,n))
+   ss[i,]=ht
+ }
```

Boxplots of the estimators:

```
> colnames(ss) <-
+ c("poisson","rsyst","rpivotal","tille","midzuno","syst","pivotal","srswor")
> boxplot(data.frame(ss), las=3)
>
```

