Unequal probability sampling designs

July 10, 2025

1 Examples of maximum entropy sampling design and related functions

a) Example 1

Consider the Belgian municipalities data set as population, and a sample size n=50

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

Compute the inclusion probabilties proportional to the 'averageincome' variable

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

Draw a random sample using the maximum entropy sampling design

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

The sample is

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

Compute the joint inclusion probabilities

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

Check the result

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

b) Example 2

Selection of samples from Belgian municipalities data set, sample size 50. Once the matrix q (see below) is computed, a sample is quickly selected. Monte Carlo simulation can be used to compare the true inclusion probabilities with the estimated ones.

```
> 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)
```

Monte Carlo simulation to check the sample selection; the difference between pik and the estimated inclusion prob. (object tt below) 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 Example of unequal probability (UP) sampling designs

Selection of samples from 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 using auxiliary information in sampling designs. We use:

- some πps sampling designs with Horvitz-Thompson estimation, using auxiliary information in a sampling desing (size measurements of population units in 2004);
- 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:

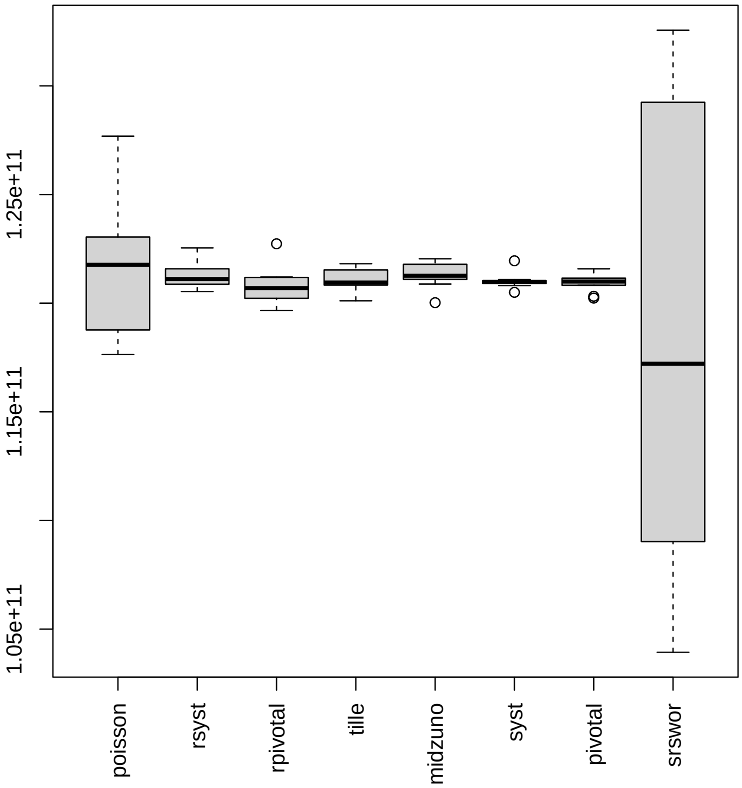
```
> y=belgianmunicipalities$TaxableIncome
```

Simulation and computation of the Horvitz-Thompson estimators:

```
> 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)
>
```



1.05e+11
1.15e+11
1.25e+11
poisson
rsyst
rpivotal
tille
midzuno
syst
pivotal
srswor