

# Unequal probability sampling designs

January 21, 2011

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  $\pi$  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)

```

