



Developments on Coordinated Poisson Sampling

Lionel Qualité

Office Fédéral de la Statistique

14/09/2011



Sample Coordination

- ▶ Control the overlap between survey samples,
- ▶ spread the burden as evenly as possible on the population (negative coordination),
- ▶ hopefully get better response rates,
- ▶ or estimate more precisely evolutions and differences (positive coordination),
- ▶ does not reduce the response burden.



Coordinated Poisson Sampling

- ▶ Extends Brewer et al. (1972)'s method of two samples selection with permanent random numbers,
- ▶ allows to select coordinated one-occasion surveys, panels or rotating panels,
- ▶ accomodates dynamic populations with births, deaths, mergers or splits,
- ▶ has transversal Poisson sampling designs (independent unequal-probabilities unit selections),
- ▶ has some optimality property for sample coordination.



Situation - notations

- ▶ Population on different occasions t , U^t , $t = 1, \dots$,
- ▶ inclusion probability of unit k in t^{th} sample : π_k^t ,
- ▶ joint inclusion probabilities at times t and s : π_k^{ts} ,
- ▶ independent surveys : $\pi_k^{ts} = \pi_k^t \pi_k^s$,
- ▶ positive coordination for unit k if $\pi_k^{ts} > \pi_k^t \pi_k^s$, negative otherwise,
- ▶ “optimal” coordination at bounds
$$\max(0, \pi_k^t + \pi_k^s - 1) \leq \pi_k^{ts} \leq \min(\pi_k^t, \pi_k^s).$$

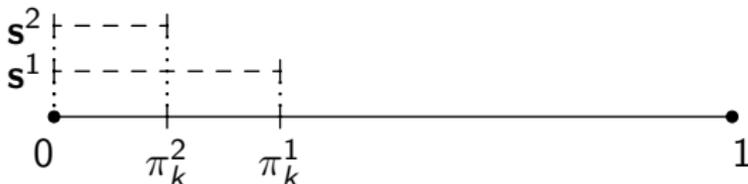


Brewer's two samples selection method - 1

- ▶ First sampling occasion



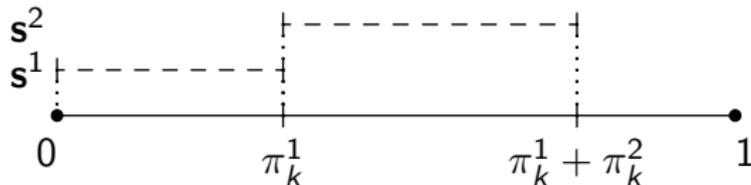
- ▶ Positive coordination when $\pi_k^2 \leq \pi_k^1$



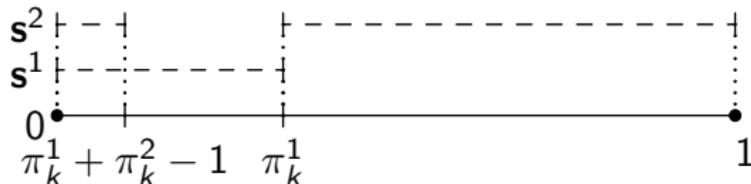


Brewer's two samples selection method - 2

- ▶ Negative coordination when $\pi_k^1 + \pi_k^2 \leq 1$



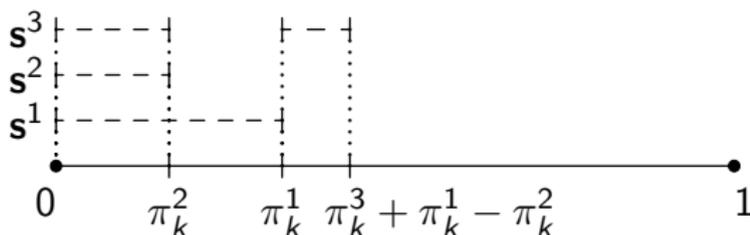
- ▶ Negative coordination when $\pi_k^1 + \pi_k^2 \geq 1$





Generalization to 3 or more surveys

- 1 Put an order on sub-intervals of $[0, 1]$ according to desired coordination rules,
- 2 construct selection zone for new survey,
- 3 example : third survey positively coordinated with second then negatively coordinated with first.





Implementation

- ▶ Order of coordination priority required,
- ▶ linear growth of datasets and computation times,
- ▶ in use since 10/2009 for business surveys and 11/2010 for population surveys,
- ▶ 530'000 units selected from a population of 8'000'000, in 17 surveys : no multiple selections as opposed to 8'200 expected under independent sampling



Problematic aspects

- 1 Random sample size
 - ▶ uncontrolled budget
 - ▶ increased variance
- 2 Unit selection independence
 - ▶ no “multi-level” coordination e.g. household surveys and population surveys,
 - ▶ possible multiple selections within domains (such as households).



Random sample size - 1

- ▶ Usually very small impact on the overall sample size,
- ▶ but could be problematic for small surveys and prevent to control interviewers workloads
- ▶ variance of calibrated (Hájek) estimator with Poisson sampling is close to that of Horvitz-Thompson estimator with fixed size sampling (*when inclusion probabilities are equal and according to available approximations*).



Random sample size - 2

- ▶ Still a problem with small domains where we would like to have a minimum number of units,
- ▶ not a new problem : was already present due to non-response,
- ▶ solution : estimate the probability of obtaining too small samples within domains and increase inclusion probabilities in order to limit this risk,
- ▶ cost can be high if we want to have a really small risk, and sample allocation can become ineffective.



Unit independence

- ▶ No good solution for multi-level coordination,
- ▶ easiest one would be coordination at the highest level and use of multi-level sampling designs,
- ▶ but not suited to the selection of panels at lower levels,
- ▶ decided to focus on business surveys and not on local units surveys.



Effect on burden - 1

Size	Total	Number of selections (actual/exp. independent/forced)			
		1	2	3	4
0-2	305344	54/71/12	6/5/2	0/0/0	0/0/0
3-9	147076	9335/9026/473	30/138/2	1/1/0	0/0/0
10-19	29231	2905/2866/552	27/64/6	1/1/1	0/0/0
20-49	16305	5220/4582/2195	1327/1536/354	6/67/1	0/0/0
50-99	5494	1905/1914/2037	1900/1901/1992	270/263/7	0/0/0
100+	4959	1142/1126/1366	2388/2399/1941	417/415/379	1/1/1
Total	508409	20561/19585/6635	5678/6043/4297	695/746/388	1/1/1

Table: Number of selections after 4 surveys : actual/expected under independent sampling/forced.



Effect on burden - 2

Size	Number of selections (actual/exp. independent)					...
	1	2	3	4	5	
0-2	54/71	6/5	0/0	0/0	0/0	...
3-9	7605/12610	4794/2064	60/142	1/3	0/0	...
10-19	5488/5982	1532/1194	225/274	8/37	0/1	...
20-49	4324/4227	2808/2422	1033/1148	299/401	71/94	...
50-99	245/232	1682/1700	1597/1607	687/655	103/111	...
100+	409/351	834/810	952/1024	1170/1133	717/723	...
Total	18125/23472	11656/8195	3867/4195	2165/2230	891/929	...

Table: Number of selections after 7 surveys : actual/expected under independent sampling.



Other Benefits

- ▶ Easy and correct procedure for “complex” longitudinal surveys,
- ▶ one frame for all surveys,
- ▶ all surveys recorded in one place,
- ▶ uniformization of procedures.



Brewer, K., Early, L., and Joyce, S. (1972).
Selecting several samples from a single population.
Australian Journal of Statistics, 3 :231–239.



Grab, E. & Savage, I. (1954).
Tables of the expected value of $1/x$ for positive bernoulli and
poisson variables.
Journal of the American Statistical Association **49**, 169–177.



Marciniak, E. & Wesolowski, J. (1999).
Asymptotic eulerian expansions for binomial and negative binomial
reciprocals.
Proceedings of the American Mathematical Society **127**, 3329–3338.



Qualité, L. (2009).
Unequal probability sampling and repeated surveys.
Thèse de doctorat, Université de Neuchâtel, Neuchâtel, Suisse.



Thionet, P. (1963).
Sur le moment d'ordre (-1) de la distribution tronquée. application à
l'échantillonnage de hájek.
Publ. Inst. Statist. Univ. Paris 12 **31** :827, 93–102.