Propriétés asymptotiques de la distribution d'un échantillon dans le cas d'un plan de sondage informatif

Daniel Bonnéry

N° d'ordre : 4465 ANNÉE 2011

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Mathématiques et applications
Ecole doctorale Matisse

présentée par

Daniel Bonnéry

préparée à l'unité de recherche UMR 6625 IRMAR
Institut de Recherche Mathématique de Rennes
Composante universitaire : Ensai

Propriétés asymptotiques de la distribution d'un échantillon dans le cas d'un plan de sondage informatif

Thèse soutenue à l'Ensai
le 24 novembre 2011

devant le jury composé de :

Jay BREIDT
Professeur, Colorado Sate University / co-directeur de thèse

Hervé CARDOT
Professeur, Université de Bourgogne/rapporteur

François COQUET
Professeur, Ensai/co-directeur de thèse

Jean-Claude DEVILLE
Inspecteur général de l'Insee, Ensai/examinateur

Chris SKINNER
Professeur, London School of Economics and Political Science/rapporteur

Contents

Remerciements v

Abstract vii

Résumé vii

1 Introduction 1
English version . . . 1
Version française . . . 4
References . . . 14

2 Informative selection and sample distribution 15
2.1 Population, samples, design measures and inclusion probabilities . . . 15
2.2 Study variables and design variables . . . 16
2.3 Population model . . . 19
2.3.1 Parametric fixed population model for design-based inference . . . 19
2.3.2 The iid superpopulation model for model-based inference . . . 20
2.3.3 General case: exchangeable population model . . . 20
2.4 Informative selection . . . 21
2.4.1 Informative selection in the general case . . . 22
2.4.2 Informative selection under the iid superpopulation model . . . 23
2.4.3 Informative selection under the fixed population model . . . 23
2.5 Asymptotic framework . . . 24
2.6 The sample pdf and the limit sample pdf . . . 25
References . . . 27

3 Uniform convergence of the sample cdf 29
3.1 Results . . . 29
3.1.1 Asymptotic framework and assumptions . . . 29
3.1.2 Uniform convergence of the empirical sample cdf . . . 32
3.2 Examples . . . 34
3.2.1 Non-informative selection without replacement . . . 34
3.2.2 Length-biased sampling . . . 35
3.2.3 Cluster sampling . . . 36
3.2.4 Cut-off sampling and take-all strata . . . 39
3.2.5 With-replacement sampling with probability proportional to size . . . 40
3.2.6 Endogenous stratification . . . 43

3.3 Conclusion . . . 45
References . . . 45

4 Kernel density estimation 47
4.1 Definitions . . . 47
4.2 Properties of the kernel estimator . . . 48
4.3 Link to the Horvitz Thompson estimator of the sample cdf . . . 55
4.4 Examples . . . 55
4.4.1 Non-informative selection . . . 55
4.4.2 Cluster sampling . . . 56
4.4.3 With-replacement sampling with probability proportional to size . . . 57
References . . . 58

5 Sample likelihood estimation 59
5.1 Notations and definitions . . . 59
5.1.1 Assumptions . . . 59
5.1.2 Further definitions . . . 60
5.1.3 Assumptions on the design measure: asymptotic independence of draws . . . 60
5.2 Maximum sample likelihood estimation . . . 60
5.2.1 Approximation of log-likelihood based on the sample distribution . . . 61
5.2.2 Maximum likelihood estimation based on the sample distribution . . . 61
5.3 Example: stratified sampling . . . 62
5.3.1 Asymptotic framework . . . 62
5.3.2 Maximum sample likelihood estimator . . . 63
5.3.3 Existence of a consistent root to the MLE equation and limiting variance . . . 64
5.3.4 Comparison to other estimators . . . 65
5.3.5 Simulations . . . 65
References . . . 66

6 Optimal inclusion probabilities for balanced sampling 69
6.1 Introduction . . . 69
6.2 Notation and balanced sampling . . . 70
6.3 Optimal allocation for balanced sampling . . . 72
6.3.1 Optimal allocation for an approximation of the variance . . . 72
6.3.2 Generalization of the approximated optimization problem . . . 73
6.3.3 Practical implementation of the optimization problem . . . 75
6.4 A simulation study . . . 77
6.4.1 Simulation 1: optimal inclusion probabilities . . . 77
6.4.2 Simulation 2: approximately optimal inclusion probabilities . . . 79
6.5 Optimal inclusion probabilities for probabilistic quota sampling . . . 82
6.5.1 Probabilistic quota sampling . . . 82
6.5.2 Optimal inclusion probabilities . . . 82
6.6 Concluding remarks . . . 84
References . . . 84

Conclusion 87

A Essential mathematical notation 89
A.1 Set theoretic notation and terminology . . . 89
A.2 Derivation . . . 89
A.3 Measure and probability . . . 89
A.4 Algebra . . . 90
A.5 Miscellaneous . . . 90

B Proofs for chapter 2 91
B.1 Proof of Property 2.7 . . . 91

C Proofs for chapter 3 93
C.1 Proofs of Theorems 3.1 and 3.2 . . . 93
C.1.1 Proof of Theorem 3.1: uniform L_2 convergence of the empirical cdf . . . 93
C.1.2 Proof of Theorem 3.2: uniform almost sure convergence of the empirical cdf . . . 96
C.2 Proof of Corollaries 3.1 and 3.2 . . . 99
C.2.1 Proof of Corollary 3.1 . . . 99
C.2.2 Proof of Corollary 3.2 . . . 100
C.3 Proofs for specific designs . . . 100
C.3.1 Proof of A1 in the case of sampling with replacement . . . 100
C.3.2 Proof for stratified simple random sampling without replacement, with non random number of strata stratum sizes, and stratum sample sizes . . . 105
References . . . 108

D Proofs for chapter 5 109
D.1 Proof of Lemma 5.1 . . . 109
D.2 Proof of Theorem 5.1 . . . 110
D.3 Proof of Theorem 5.2 . . . 111
D.4 Proofs for stratified sampling . . . 112
D.4.1 Proof of Result 5.1 . . . 112
D.4.2 Proof of Result 5.3 . . . 113
D.4.3 Asymptotic normality . . . 115
D.4.4 Proof of Result 5.2 . . . 120
D.4.5 Proof of Result 5.4 . . . 121
D.4.6 Proof of Result 5.5 . . . 121
References . . . 121

E Proofs for chapter 6 123
E.1 Proof of equation . . . 123
E.2 Proof of Theorem 6.1 . . . 124

F R-code used for simulations 125

General Bibliography 141
Index of terms . . . 145
Index of notations . . . 146
List of figures . . . 148
List of Tables . . . 148
List of Algorithms . . . 148

Remerciements

Je tiens à remercier Jay Breidt et François Coquet, pour la confiance qu'ils m'ont accordée en me permettant de débuter une thèse, leur patience, et l'aide précieuse et indéfectible qu'ils m'ont apportée pour que je puisse, avec eux, mener à terme ce projet. Je remercie vivement Jay Breidt de m'avoir accueilli à l'université d'état du Colorado, et ainsi permi de vivre une expérience très enrichissante.

Durant les trois années à l'Ensai, j'ai pu compter sur les conseils et le soutien de Guillaume Chauvet, Jean-Claude Deville, Éric Lesage et Valentin Patilea, et je les en remercie.

Je suis très honoré de l'acceptation de Chris Skinner et Hervé Cardot d'être rapporteurs de cette thèse.

Je suis honoré et reconnaissant à Jean-Claude Deville d'avoir accepté d'être membre de mon jury de thèse.

Je salue au passage Myriam Vimond, Samuel Balmand, Samuel Maistre, Brigitte Gelein, Laurent Rouvière, Jocelyn Julienne, Pierre Clauss, Magalie Fromont et Marian Hristache avec qui j'ai passé d'agréables moments à l'Ensai.

Abstract

Consider informative selection of a sample from a finite population. Responses are realized as independent and identically distributed (iid) random variables with a probability density function (pdf) f, referred to as the superpopulation model. The selection is informative in the sense that the sample responses, given that they were selected, are not iid f. A limit sample pdf is defined, which corresponds to the limit distribution of the response of a unit given it was selected, when population and sample sizes grow to ∞. It is a weighted version $\rho.f$ of the population pdf. In general, the informative selection mechanism may induce dependence among the selected observations. The impact of the dependence among the selected observations on the behavior of basic distribution estimators, the (unweighted) empirical cumulative distribution function (cdf) and the kernel density estimator of the pdf, is studied. An asymptotic framework and weak conditions on the informative selection mechanism are developed under which these statistics computed on sample responses behave as if they were computed from an iid sample of observations from $\rho.f$. In particular, the empirical cdf converges uniformly, in L_2 and almost surely, to the corresponding version of the superpopulation cdf, yielding an analogue of the Glivenko-Cantelli theorem. Further, we compute the rate of convergence of the kernel density estimator to the limit sample pdf. When weak conditions on the selection are satisfied, one can consider that the responses are iid $\rho.f$ in order to make inference on the population distribution. For example, if the response pdf belongs to a parametrized set $\{f_\theta\}$, and the stochastic dependence between the design and response variables is well known, then the likelihood derived as the product of limit sample pdf's can be used to compute a maximum sample likelihood estimator of θ. Convergence and asymptotic normality of this estimator is established.

The last part of the dissertation deals with balanced sampling. Consider a sampling design balanced on a set of design variables z, which may depend on the inclusion probabilities. The variance of the Horvitz-Thompson estimator of the total of a study variable y can be approximated by a function of y, z, and the inclusion probabilities. We propose algorithms that compute the inclusion probabilities that minimize this approximate variance.

Résumé

Considérons la sélection d'un échantillon d'une population finie selon un plan de sondage informatif, et le modèle de superpopulation suivant : à chaque élément de la population, correspond la réalisation d'une variable aléatoire, les réalisations sur la population sont supposées indépendantes et idendiquement distribuées (iid) selon une loi qui admet une densité f par rapport à la mesure de Lebesgue. Le plan de sondage est informatif dans le sens où le vecteur des réalisations qui correspondent aux éléments de l'échantillon n'est pas un vecteur de variables aléatoires indépendantes, et la loi d'une réalisation conditionnelle à la sélection de l'élément correspondant n'est pas égale à la loi initiale des réalisations sur la population. Une loi de probabilité limite et une densité de probabilité limite des réalisations sur l'échantillon sont définies ; elles correspondent à la limite de la distribution d'une réalisation sur l'échantillon lorsque les tailles de la population et de l'échantillon tendent vers l'infini. La densité de la distribution limite de l'échantillon est

une version pondérée, notée $\rho.f$, de la densité initiale f. En général, le processus aléatoire de sélection peut induire une dépendance entre les réalisations correspondant aux éléments sélectionnés. L'impact d'une telle dépendance sur le comportement asymptotique d'estimateurs classiques, la fonction de répartition empirique et un estimateur à noyau de la densité, est étudié. Un cadre asymptotique et des conditions faibles sur le processus de sélection sont donnés, sous lesquels ces statistiques ont les mêmes propriétés asymptotiques que les mêmes statistiques calculées à partir d'un vecteur de variables iid et de densité $\rho.f$ par rapport à la mesure de Lebesgue. En particulier, la fonction de répartition empirique converge uniformément, dans L^2 et presque sûrement, vers la version pondérée de la fonction de répartition des observations de la population, ce qui constitue un résultat analogue au théorème de Glivenko-Cantelli. Par ailleurs, nous donnons la vitesse de convergence de l'estimateur à noyau de la densité vers la densité limite de l'échantillon. Quand des conditions faibles sur le processus de sélection sont vérifiées, ces résultats sont des premières indications selon lesquelles il est possible de considérer que les réalisations sur l'échantillon sont iid et de densité $\rho.f$ approximativement, notamment dans une perspective d'inférence sur f. Par exemple, étant donné un modèle paramétrique $f \in \{f_\theta\}_{\theta\in\Theta}$ pour la loi des réalisations sur la population, lorsque le lien de dépendance stochastique entre le processus de sélection et les réalisations sur la population est connu, alors la vraisemblance approchée de l'échantillon, définie comme produit de densités limites, peut être utilisée pour calculer un estimateur de maximum de vraisemblance de θ. La convergence et la normalité asymptotique d'un tel estimateur sont établies.

La dernière partie de la thèse traite de tirage équilibré. Etant donné un sondage équilibré sur un jeu de variables auxiliaires z, qui peuvent dépendre des probabilités d'inclusion, la variance de l'estimateur de Horvitz-Thompson du total d'une variable d'intérêt y peut être approchée par une fonction de y, de z et des probabilités d'inclusion. Des algorithmes de calcul de probabilités d'inclusion fonctions de y et de z qui minimisent cette variance approchée sont proposés, puis adaptés au cas où seule une information partielle sur y et z est disponible.

Chapter 1

Introduction

English version

The aim of survey sampling is to draw conclusions about the characteristics of a complete finite population from data corresponding to a sample from the population. For a statistician, drawing conclusions consists of making statistical inference. In survey sampling, two kinds of inference are usually contrasted. In model-based inference, the characteristics of the elements of the population are considered random, and the aim of inference is the random process that generates those characteristics. This random process is known as a superpopulation model. In design-based inference, the characteristics are considered fixed and constitute the target of the inference, and only the randomness of the sample is used in making inference. This inferential structure is then referred to as the fixed population model. Defining a general statistical model that is suitable for both model-based and design-based inference is possible, but unfortunately, distinctions must often be made when stating definitions and results.

Before we introduce the concept of informative selection, we define the basic concepts of survey sampling: a finite population is defined as a finite set, composed of elements. Without loss of generality, let $U = \{1 \dots N\}$ denote the finite population, with N the population size. A sample is a subset of U.

To any element in the population, are associated some characteristics. Some characteristics are of interest to the analyst, and are called response variables or study variables. For example, gender and income could be response variables in the case of a survey concerning a population of people. The analyst does not observe the characteristics of all elements in the population, but only for the elements of a sample. This sample is selected according to a random process, called the sampling design, constructed by those responsible for the survey. In this construction, the survey designer may use some characteristics known for all elements of the population. These characteristics are called design variables. For example, postal code might be known for all households in a survey of human population.

In noninformative selection (e.g., Cassel et al. (1977, §1.4) or Särndal et al. (1992, Remark 2.4.4)), the probability of drawing the sample does not depend explicitly on the study variables. If instead there is stochastic dependence between the design variables and the study variables, then the probability law of a study variable in the sample can be different from the law of the same study variable in the population. Specifically, suppose the response variables Y follow the superpopulation model in which they are realized as iid random variables with probability density function (pdf) f. The selection is informative in the sense that the sample responses, *given that they were selected*, are not iid f (a specification of informative selection that includes the iid case described here is given in Pfeffermann and Sverchkov (2009, Remark 1.2)). Under informative selection, the selection process has to be taken into account when making inference. Consider for example a survey in which women are always overrepresented: if the aim of the survey is inference on the distribution of the incomes in the entire population, then the selection process is informative, and

must be considered. Informative selection may also induce dependence among the selected observations. Consider for example a selection mechanism balanced on a proportion of men equal to 50% in the sample. Observations cannot be considered independent and identically distributed (iid) because of the constraint introduced by the selection mechanism.

Nevertheless, a large body of current methodological literature treats the observations as if they were independently distributed according to the *sample pdf*, defined as the conditional distribution of the random variable Y, given that it was selected (under informative selection, the sample pdf differs from f). In particular, Pfeffermann et al. (1998) (see some motivating work in Skinner, 1994) have developed a *sample likelihood* approach to estimation and inference for the superpopulation model, which maximizes the criterion function formed by taking the product of the sample pdf's, as if the responses were iid. This methodology has been extended in a number of directions (Eideh and Nathan, 2006, 2007, 2009; Pfeffermann et al., 2006; Pfeffermann and Sverchkov, 1999, 2003, 2007). An extensive review of these and other approaches to inference under informative selection is given by Pfeffermann and Sverchkov (2009).

The aim of this dissertation is to study the implications of informative selection for inference on the population model. More particularly we show that under some conditions on the sampling design, we can treat observations on the sample as if they were independent and identically distributed, according to a weighted version of their initial distribution.

The purpose of chapter 2 is the description of a general asymptotic model for survey sampling in the case of informative selection. In chapter 2, basic notions on survey sampling are given, and the general model (including both the superpopulation model and the fixed population model) for responses observed on the sample is precisely described. Though the methodology that uses a sample likelihood is mainly used in the context of model-based inference, an effort has been made for the general model to embrace both the fixed population model used for design-based inference, and the superpopulation model, used for model-based inference. Our framework also allows treatment of both with and without replacement sampling. A definition of informative selection is proposed for design-based and model-based inference. In the model-based approach, the definition of the sample pdf is generalized by taking into account sampling without replacement and random sample size. A definition of the limit sample pdf is also proposed. The limit sample pdf is a weighted version, $\rho.f$, of the pdf of the response value.

We consider the problem of identifying a suitable model using observed data under informative selection. In an ordinary inference problem with iid observations, we often begin not by constructing a likelihood and conducting inference, but by using basic sample statistics to help identify a suitable model. In particular, under iid sampling the empirical cumulative distribution function (cdf) converges uniformly almost surely to the population cdf, by the Glivenko-Cantelli theorem (e.g., van der Vaart, 1998, Theorem 19.1). What is the behavior of the empirical cdf under informative selection from a finite population? In chapter 3, under the general asymptotic framework presented in chapter 2, we propose weak conditions on the informative selection mechanism under which the (unweighted) sample empirical cdf converges uniformly, in L_2 and almost surely, to the weighted version of the superpopulation cdf (the cdf corresponding to the pdf $\rho.f$); that is, the empirical cdf behaves as if the observations were iid $\rho.f$. The convergence results are given for both the fixed population model and the superpopulation model. The corresponding quantiles also converge uniformly on compact sets. Our almost sure results rely on an embedding argument. Importantly, our construction preserves the original response vector for the finite population, not some independent replicate.

The conditions we propose are verifiable for specified designs, and involve computing conditional versions of first and second-order inclusion probabilities. Motivated by real problems in surveys and other observational studies, we give examples of where these conditions hold and where they fail. Where the conditions hold, the convergence results we obtain may be useful in making inference about the superpopulation model. For example, the results may be used in identifying a suitable parametric family for the weighted cdf, from which a selection mechanism and a superpopulation pdf may be postulated using results

in Pfeffermann et al. (1998). The work presented in chapter 3 has been accepted for publication in *Bernoulli* (see Bonnéry et al. (2011)).

Our results in chapter 4 continue the theme of non-parametric estimation of distributional properties, this time via kernel density estimation. Under the asymptotic framework described in chapter 2, verifiable conditions are given under which another class of non-parametric distribution estimators, the kernel density estimators, behave as if the responses were iid $\rho.f$. We study the rate of convergence of such statistics to the limit sample pdf. When the weighting function ρ is available, this offers opportunities to get alternative estimators to survey-weighted estimators that use Horvitz-Thompson plug-in methods.

In chapter 5, we consider a parametric version of the superpopulation model described in chapter 2. Assume the response pdf is f_θ, and assume the law of the design variable conditional on the response is indexed by a parameter ξ. The aim of the inference is the estimation of θ, the parameter that corresponds to the law of the response values, and we assume we have a consistent estimator $\hat{\xi}$ of ξ. This estimator is plugged into the approximation of the sample likelihood and we study the properties of the estimator $\hat{\theta}$ that maximizes this criterion in θ. Adapting Gong and Samaniego (1981), we prove the existence, consistency and rate of convergence to a normal distribution of this estimator. As in chapter 3 and 4, the conditions we propose are verifiable for a list of sampling designs, commonly encountered in real problems in surveys. Under a strong set of assumptions (in particular, sample size remains fixed as population size goes to infinity), Pfeffermann et al. (1998) have established the pointwise convergence of the joint distribution of the responses to the product of the sample pdf's. This is taken as partial justification of the sample likelihood approach. Alternatively, the full likelihood of the data (including selection indicators for the finite population and response variables and inclusion probabilities for the sample) can be maximized (Breckling et al., 1994; Chambers et al., 1998), or the *pseudo-likelihood* can be obtained by plugging in Horvitz-Thompson estimators for unknown quantities in the log-likelihood for the entire finite population (e.g. Binder, 1983; Chambers and Skinner, 2003; Kish and Frankel, 1974, §2.4). We show that using the likelihood derived as the product of limit sample pdf's is possible and can allow estimators with smaller variance than the pseudo-likelihood estimators derived from Horvitz-Thompson plug-in methods. We illustrate this final result by simulations applied to stratified sampling, for which we prove that the stated assumptions hold. This leads us to prove a central limit theorem for estimators from the approximate likelihood in the particular case of stratified sampling with fixed number of strata.

In chapter 6, results on balanced sampling (see Deville and Tillé (2004, 2005)) are presented. This work is in collaboration with G. Chauvet and J.C. Deville. The results consist of the properties of two algorithms that allow computation of optimum inclusion probabilities. Deville and Tillé (2005) propose an approximate variance of the Horvitz-Thompson estimator of the total of a study variable y under balanced sampling. Consider a sampling design balanced on a set of design variables z, which may depend on the inclusion probabilities. The approximate variance is a function of y, z, and the inclusion probabilities. The optimum inclusion probabilities minimize this approximate variance. The first algorithm is proposed for a specific case where the balancing characteristics are qualitative and do not depend on the inclusion probabilities. The existence of a global solution to the optimization problem and the convergence of the algorithm to this solution are proved. The second algorithm is proposed for a case where the balancing variables do depend on the inclusion probabilities. The convergence to a local solution of the minimization problem is proved. For the first algorithm, simulations are made, which show that in some cases, minimizing the approximation of variance leads to a reduction of the true variance, as estimated by Monte Carlo. Part of this work has been published in the *Journal of Statistical Planning and Inference* (Chauvet et al., 2011). Previously, a more extended version, was published in the conference proceedings, *Actes des Journées de Méthodologie Statistique of the Insee* (Bonnéry et al., 2009).

Long proofs are given in the appendices. Appendix A is dedicated to some necessary mathematical background and notation used in the document. An index of notations and an index of terms are provided

starting on p. 146.

Version française

Le but d'une enquête par sondage est de tirer des conclusions sur des caractéristiques d'une population finie à partir de données observées sur un échantillon de cette population. Pour un statisticien, tirer des conclusions revient à faire de l'inférence statistique. En théorie des sondages, deux types d'inférence sont habituellement présentées. Pour l'inférence basée sur le modèle, les caractéristiques des éléments de la population sont considérées comme aléatoires, et la cible de l'inférence est le processus aléatoire qui génère ces caractéristiques. Les hypothèses faites sur ce processus aléatoire avant la conduite de l'inférence correspondent à un modèle de superpopulation. Pour l'inférence basée sur le plan, les caractéristiques sont considérées comme fixes et constituent la cible de l'inférence, et le seul processus aléatoire qui entre en jeu est la sélection de l'échantillon. De ce fait, l'inférence est basée sur l'aléa introduit par la sélection de l'échantillon. Le modèle statistique correspondant est appelé modèle de population fixe. Définir un modèle général qui contient les modèles de superpopulation et de population fixe est possible, mais l'énoncé de définitions ou de résultats nécessite souvent une distinction, pour plus de pertinence.

Avant d'introduire le concept de sélection informative, nous définissons les concepts fondamentaux de la théorie des sondages : une population finie est définie comme un ensemble fini U, composé d'éléments. Soit $N \in \mathbb{N}$ une taille de population et soit U une population de taille N. Sans perte de généralité, on considère que $U = \{1 \dots N\}$. Un échantillon est un sous ensemble de U.

À un élément quelconque de la population est associé une liste de caractéristiques. Certaines caractéristiques sont des caractéristiquess d'intérêt pour l'analyste et sont appelées variables d'intérêt. Par exemple, le sexe et le revenu pourraient être des variables d'intérêt pour une enquête sur une population humaine. L'analyste n'observe pas les caractéristiques de tous les éléments dans la population, mais seulement pour quelques éléments d'un échantillon. Cet échantillon est sélectionné selon un processus aléatoire appelé plan de sondage, construit par le responsable de l'enquête. Dans cette construction, le responsable de l'enquête peut être amené à utiliser quelques caractéristiques connues pour chaque élément de la population. Ces variables seront appelées variables auxiliaires.

Dans le cas de sélection non-informative, (e.g., Cassel et al. (1977, §1.4) or Särndal et al. (1992, Remarque 2.4.4)), la probabilité de sélectionner un échantillon ne dépend pas explicitement des variables d'intérêt. Si en revanche, il existe une dépendance entre les variables d'intérêt et les variables auxiliaires utilisées pour la construction du plan, alors la loi de probabilité d'une variable d'intérêt pour un individu de l'échantillon peut être différente de la loi de probabilité de cette même variable d'intérêt pour un individu de la population. Plus spécifiquement, supposons qu'une variable d'intérêt Y suit le modèle de superpopulation selon lequel les valeurs de Y sont des variables aléatoires indépendantes et identiquement distribuées de densité f par rapport à la mesure de Lebesgue. La densité f est appelée densité sur la population. La sélection est informative dans le sens ou les valeurs des variables d'intérêt pour les individus de l'échantillon ne sont pas des réalisations iid d'une loi de densité f (voir Pfeffermann and Sverchkov (2009, Remarque 1.2)). Dans le cas d'une sélection informative, le processus de sélection doit donc être pris en compte lors de l'inférence. Considérons un plan de sondage qui privilégie les échantillons où les femmes sont surreprésentées. Si le but de l'inférence est la distribution des revenus dans la population, alors le processus de sélection est informatif, et doit être pris en considération. Un processus de sélection informatif peut aussi induire une dépendance entre les valeurs observées sur l'échantillon. Par exemple, dans le cas d'un plan de sondage équilibré sur une proportion d'hommes enquêtés, les observations sur l'échantillon ne peuvent être considérées comme indépendantes et identiquement distribuées, du fait de la contrainte imposée par le mécanisme de sélection.

Cependant, certaines méthodes traitent les observations sur l'échantillon comme si elles étaient indépendantes, et identiquement distribuées selon la densité de l'échantillon, définie comme la densité conditionnelle de la $\mathrm{variable}$ aléatoire Y, conditionnellement à sa sélection. Dans le cas de sélection informative,

la densité de l'échantillon peut être différente de f. En particulier, Pfeffermann et al. (1998) (voir aussi Skinner, 1994) ont développé une méthode d'estimation et d'inférence dans le cadre du modèle de superpopulation, basée sur une approximation de la vraisemblance des observations sur l'échantillon. La vraisemblance $\mathrm{approximée}$ est le produit des valeurs des densités d'échantillon prises pour les valeurs observées sur l'échantillon, comme si les observations sur l'échantillon étaient iid. Cette méthodologie a été étendue dans plusieurs directions (Eideh and Nathan, 2006, 2007, 2009; Pfeffermann et al., 2006; Pfeffermann and Sverchkov, 1999, 2003, 2007). Pour une revue des méthodes d'inférence dans le cas de sélection informative, on pourra se référer à Pfeffermann and Sverchkov (2009).

L'objet de ce mémoire est d'étudier les implications d'une sélection informative dans la perspective d'une inférence statistique sur le modèle de superpopulation. Plus précisément, nous montrons que sous certaines conditions sur le plan de sondage, les observations sur l'échantillon peuvent être traitées comme indépendantes, et identiquement distribuées selon une version pondérée de leur distribution sur la population.

Un des objets du chapitre 2 est la description d'un modèle statistique asymptotique général pour l'inférence sur données d'enquête dans le cas de sélection informative. Pour celà, nous considérons une suite de populations $(U_\gamma)_{\gamma\in\mathbb{N}}$, et une suite de plans de sondages et d'échantillons. Nous notons N_γ la taille de la population U_γ, et $I_{\gamma k}$ le nombre d'occurences de l'élément k de U_γ dans le k^{e} échantillon tiré. $Y_{\gamma k}$ est le vecteur des variables d'intérêt associé à l'élément k de la population U_γ. Nous supposons que le numéro (k) de chaque individu n'apporte aucune information, ni même après sélection d'un échantillon, ce que nous traduisons par une hypothèse d'échangeabilité en ligne des vecteurs des variables d'intérêt sur la population et du vecteur des nombres d'occurence dans l'échantillon. Dans le chapitre 2, des notions fondamentales sont données, et le modèle général (qui inclut à la fois le modèle de superpopulation et le modèle de population fixe) est décrit précisément. Une fois défini, ce cadre asymptotique permet de traiter à la fois la sélection avec et sans remise.

Une définition de la sélection informative est proposée, valable pour l'inférence basée sur le plan et l'inférence basée sur le modèle. Dans le cas de l'inférence basée sur le plan, cette définition ne correspond pas à la définition de référence dans le cas du modèle fixe de population (Cassel et al. (1977, §1.4)). Nous soulignons que cette dernière définition est ambigüe, car elle traite de dépendance de variables non aléatoires, (les variables utilisées pour définir le plan et les variables d'intérêt), sans préciser clairement de quelle dépendance il s'agit. Il ne peut s'agir de dépendance stochastique, il doit donc s'agir de dépendance fonctionnelle, ce qui aurait nécessité une description détaillée de l'espace des paramêtres de façon à faire apparaître une correspondance fonctionnelle entre les variables d'intérêt et les variables utilisées pour le plan. L'ambiguïté de cette définition apparaît avec des exemples choisis, pour lesquels le modèle paramétrique du modèle fixe de population est donné de façon détaillée, avec variables d'intérêt et variables utilisées pour le plan vues comme paramêtres du modèle global. Toutefois, la définition de sélection informative que nous proposons, si elle a un sens, revêt peu d'intérêt dans le cadre du modèle fixe de population. La définition que nous proposons généralise la définition de sélection informative existante dans le cas de l'inférence basée sur un modèle. Le cas le plus simple de sélection non informative est celui du modèle de superpopulation iid, combiné avec un sondage aléatoire simple de taille n : la population coonstituent un échantillon (au sens de la statistique inférentielle classique) de taille N de variables iid suivant une loi initiale, et les observations sur l'échantillon sélectionné est un échantillon de taille n de variables aléatoires suivant aussi la loi initiale. Dans ce cas, l'inférence sur la loi initiale peut donc être menée sans se soucier du processus de sélection. Un autre exemple de sélection non informative est celui ou sur la population sont définies deux variables, X et Y, liées selon un modèle de régression linéaire $\forall k \in U, Y_k = X_k\beta + \varepsilon_k$ où les résidus ε_k sont indépendants et identiquement distribués, et le vecteur des résidus est indépendant du vecteur des X_k. Si on considère un plan de sondage sans remise dépendant (au sens de la dépendance de variables aléatoires, le plan de sondage

étant vu comme une variable aléatoire à valeur dans l'espace des probabilités sur l'ensemble des échantillons) de X mais indépendant du vecteur de résidus ε, la sélection est non informative sur la loi des résidus, et est non informative sur le paramètre β, car la loi conditionnelle du vecteur des Y_k pour k appartenant à l'échantillon conditionnellement au vecteur des (X_k) pour k appartenant à l'échantillon, est, pour ce plan de sondage, indépendante (au sens de l'indépendance de deux variables aléatoires) du plan de sondage. À partir de cet exemple simple, on comprend l'utilité d'une définition qui permettrait, dans le cas de plans de sondage avec ou sans remise, à taille variable ou fixe, de déterminer, étant donné un modèle de population et une loi d'intérêt quelconques, si le plan de sondage doit être pris en compte lors de l'inférence. Pour dépasser le stade de la définition heuristique, il est nécessaire de faire apparaître, de façon un peu lourde, les différents niveaux d'aléa (génération de la population et du plan, puis conditionnellement à la réalisation du plan de sondage, tirage d'un échantillon selon le plan), pour aboutir à la proposition d'une définition générale ou le plan de sondage aléatoire simple joue un rôle de référence.

Toujours dans le chapitre 2, nous proposons une définition de la loi de distribution sur l'échantillon. Il s'agit d'une définition d'une loi de distribution, valable dans le cas de plans avec ou sans remise, à taille fixe ou variable, qui correspond, dans les cas ou celle-ci est définie, à la loi d'une observation associée à un élément tiré au hasard (avec probabilités proportionnelles au nombre d'occurences) dans l'échantillon. Dans le cas de sondage sans remise et de taille fixe, la loi de distribution sur l'échantillon correspond à la loi de la variable d'intérêt conditionnelle à l'appartenance à l'échantillon. Cette loi conditionnelle a été utilisée (Pfeffermann et al. (1998)) pour approximer, lorsque la taille de la population est grande, la loi des observations sur l'échantillon par la loi d'un vecteur de même taille que l'échantillon constitué de variables indépendantes distribuées selon cette loi conditionnelle. L'inférence est alors conduite en suivant cette approximation. Le fil directeur des premiers chapitres de la thèse est le suivant : quand peut on considérer que l'inférence basée sur cette approximation est valide ? Dans le cadre asymptotique général que nous proposons, nous définissons aussi une loi de distribution limite sur l'échantillon. Il s'agit d'une limite de la loi précédente, au sens de la convergence faible des mesures. La distribution limite sur l'échantillon est présentée comme une version pondérée, de densité $\rho_\infty.f$, de la distribution de la variable d'intérêt sur la population.

Définition 1. *Lorsque* $0 < \mathrm{E}\left[I_{\gamma 1}\right] < +\infty$*, la densité de l'échantillon est définie comme :* $\rho_\gamma f_\gamma$*, où*

$$\rho_\gamma(y) = \frac{\mathrm{E}\left[I_{\gamma k}|Y_{\gamma k} = y\right]}{\mathrm{E}\left[I_{\gamma k}\right]}$$

Définition 2. *Lorsque* $\rho_\infty \lim_{\gamma\to\infty} \rho_\gamma$ *est défini, la densité limite de l'échantillon est définie comme la densité de probabilité* $\rho_\infty.f$*.*

Considérons le problème d'identification d'un modèle à partir de données observées dans le cas d'une sélection informative. Dans le cas usuel d'échantillon de variables aléatoires indépendantes et identiquement distribuées, la première étape consiste souvent non pas à calculer une vraisemblance et à l'utiliser pour l'inférence, mais à utiliser des statistiques d'analyse descriptive pour tenter d'identifier un modèle adéquat. En particulier, dans le cas d'un échantillon iid d'une variable aléatoire, la fonction de répartition empirique converge uniformément presque sûrement vers la fonction de répartition de cette variable aléatoire, d'après le théorème de Glivenko-Cantelli (e.g., van der Vaart, 1998, Théorème 19.1). Quel est le comportement de la fonction de répartition empirique dans le cas d'une sélection informative d'un échantillon d'une population finie ? Dans le chapitre 3, sous le cadre asymptotique général du chapitre 2, dans le cas où la variable d'intérêt est une variable uni-dimensionnelle Y, les valeurs de Y sur la population étant des réalisations indépendantes et identiquement distribuées selon une loi de densité f par rapport à la mesure de Lebesgue, nous proposons des conditions faibles sur le mécanisme de sélection sous lesquelles la fonction de répartition

empirique converge uniformément, dans L_2 vers la version pondérée de la fonction de répartition (la fonction de répartition associée à la densité $\rho_\infty.f$).

Théorème 1. *Sous certaines conditions (cf. théorème 3.1, p.32), la fonction de répartition empirique sur l'échantillon, définie par*

$$F_\gamma = \left(\sum_{k=1}^{N_\gamma} I_{\gamma k}\right) n^{-1} \sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_k) I_{\gamma k}$$

vérifie :

$$\|F_\gamma - F_\infty\|_\infty \xrightarrow[\gamma\to\infty]{L_2} 0,$$

avec

$$\|F_\gamma - F_\infty\|_\infty = \sup_{\alpha\in\mathbb{R}} \{|F_\gamma(\alpha) - F_\infty(\alpha)|\},$$

et

$$F_\infty : \alpha \mapsto \int_{-\infty}^{\alpha} \rho_\infty f \, \mathrm{d}\lambda.$$

Les conditions que nous proposons sont vérifiables pour des plans de sondage spécifiques, et portent sur les versions conditionnelles des probabilités d'inclusion d'ordre 1 et 2. Ces conditions peuvent s'interpréter comme une indépendance asymptotique des tirages des éléments de l'échantillon deux à deux : nous définissons les fonctions :

$$\begin{aligned}
m_\gamma(y_1) &= \mathrm{E}\left[I_{\gamma 1}|Y_1 = y_1\right] \\
m'_\gamma(y_1, y_2) &= \mathrm{E}\left[I_{\gamma 2}|Y_1 = y_1, Y_2 = y_2\right] \\
v_\gamma(y) &= \mathrm{Var}\left[I_{\gamma 1}|Y_1 = y\right] \\
c_\gamma(y_1, y_2) &= \mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2}|Y_1 = y_1, Y_2 = y_2\right] \\
\delta_\gamma(y_1, y_2) &= m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) - m_\gamma(y_1) m_\gamma(y_2).
\end{aligned}$$

Ces fonctions sont utilisées pour contrôler la dépendance asymptotique des tirages, conditionnellement aux valeurs de la variable d'intérêt (via c_γ et δ_γ), et la variabilité de l'espérance de la probabilité de sélection (via v_γ). Les conditions que nous définissons impliquent que la taille de l'échantillon doit croître presque sûrement vers $+\infty$, et nous nous démarquons donc des travaux de Pfeffermann et al. (1998) qui, dans l'énoncé de leurs résultats, considèrent une taille de population fixe et constante (i.e. ne variant pas avec γ lorsque la taille de la population croît). Nous donnons ensuite des conditions pour obtenir la convergence presque sûre de la fonction de répartition. Une d'entre elle introduit une dépendance particulière entre les plans de sondages de la suite de plans de sondages. Un des intérêts de ce résultat est de mettre en avant la particularité de l'asymptotique en sondage : la suite des échantillons ne peut être vue comme une suite emboîtée où sont simplement ajoutées de nouvelles réalisations indépendantes d'une certaine loi, et la convergence presque sûre est moins naturelle qu'en statistique inférentielle classique. Nous obtenons le résultat suivant :

Théorème 2. *Sous certaines conditions sur la dépendance des tirages associés à la suite de plan de sondage, et sous des conditions d'indépendance asymptotique de sélection de deux individus (cf. théorème 3.2, p.33),*

$$\|F_\gamma - F_\infty\|_\infty \xrightarrow[\gamma\to\infty]{p.s.} 0.$$

Les résultats de convergence sont donnés pour à la fois le modèle de superpopulation et le modèle de population fixe. Les quantiles empiriques convergent aussi uniformément sur les ensembles compacts. Nous notons les quantiles empiriques $\xi_\gamma(x) = \inf\{\alpha|F_\gamma(\alpha) \geqslant x\}$, et les quantiles de la distribution sur l'échantillon $\xi_\infty(x) = \inf\{\alpha|F_\infty(\alpha) \geqslant x\}$.

Corollaire 1. *Si F_∞ est continue sur $\mathbb{R}$ et si $0 < F_\infty(y_1) = F_\infty(y_2) < 1 \Rightarrow y_1 = y_2$, alors, sous des hypothèses d'indépendance asymptotique de sélection de deux éléments, (cf. corollaire 3.1, p.33) les quantiles empiriques convergent uniformément sur tout compact, en probabilité, vers les quantiles de la loi limite de distribution sur l'échantillon. Quelque soit K un sous-ensemble compact de $]0,1[$,*

$$\sup_{x\in K} |\xi_\gamma(x) - \xi_\infty(x)| \xrightarrow[\gamma\to\infty]{P} 0.$$

Dans le cas où f a un support compact, la convergence est uniforme dans L_2:

$$\sup_{x\in]0,1[} |\xi_\gamma(x) - \xi_\infty(x)| \xrightarrow[\gamma\to\infty]{L_2} 0.$$

Nous considérons des exemples, qui correspondent à des problèmes de sondage recontrés en pratique, où ces conditions sont vérifiées et d'autres où elles ne le sont pas. Ces conditions sont notamment vérifiées dans des cas de plans de sondage non informatifs, tel le plan de sondage aléatoire simple, ou, de façon tout aussi triviale, tout plan de sondage de taille fixe sans remise indépendant de la variable d'intérêt. Par ailleurs, il est possible de construire une suite de plans de sondage informatifs où l'indépendance asymptotique des tirages n'est pas assurée, par exemple des plans en grappes où les grappes sont très liées à la variable d'intérêt. Lorsque les conditions sont remplies, les résultats de convergence obtenus peuvent être utiles pour l'inférence sur le modèle de superpopulation. Par exemple, les résultats peuvent être utiles pour identifier un modèle paramétrique convenable pour la fonction de répartition empirique pondérée (c'est à dire associée à la distribution de l'échantillon). À partir de ce modèle, étant donné un mécanisme de sélection, une famille paramétrique convenable pour la loi de densité associée à la distribution de l'échantillon peut être obtenue, en utilisant les résultats de Pfeffermann et al. (1998).

Le travail présenté dans le chapitre 3 a été accepté pour publication dans *Bernoulli* (voir Bonnéry et al. (2011)).

Les résultats du chapitre 4 portent aussi sur les propriétés d'estimateurs non paramétriques de la distribution. Sous le cadre asymptotique décrit dans le chapitre 2, des conditions vérifiables sont données sous lesquelles les estimateurs à noyau de la densité se comportent comme si les valeurs observées sur l'échantillon étaient des réalisations iid d'une loi de densité $\rho_\infty.f$ par rapport à la mesure de Lebesgue. Nous étudions la vitesse de convergence de telles statistiques vers la densité limite sur l'échantillon. Nous obtenons le résultat suivant :

Théorème 3. *Sous les conditions d'indépendance asymptotique de sélection de deux individus suivantes : $\exists V$ un voisinage de y_0, et v_∞ une fonction réelle mesurable continue en y_0 telle que*

$$\begin{cases} \lim\limits_{\gamma\to\infty} \sup\limits_{y\in V} |v_\gamma(y) - v_\infty(y)| = 0 \\ \sup\limits_{y\in\mathbb{R}} \{(v_\gamma f)(y)\} = O_\gamma(1), \quad \sup\limits_{y\in\mathbb{R}} \{(v_\infty f)(y)\} < +\infty \\ \sup\limits_{(y_1,y_2)\in\mathbb{R}^2} \{\delta_\gamma(y_1,y_2) + c_\gamma(y_1,y_2)\} = O_\gamma(1) \\ \sup\limits_{(y_1,y_2)\in V\times V} \{\delta_\gamma(y_1,y_2) + c_\gamma(y_1,y_2)\} = 0, \end{cases}$$

étant donné une séquence de fenêtres $(h_\gamma)_{\gamma\in\mathbb{N}}$ et un noyau K vérifiant des conditions standards, et sous des conditions de régularité standards de la suite de densités $(\rho_\gamma f)_{\gamma\in\mathbb{N}}$ (cf proposition 4.6, p.54, proposition 4.5, p.52) et de la densité limite $\rho_\infty f$, nous obtenons la convergence de l'estimateur à noyau de la densité $\rho_\infty f$:

$\forall y_0 \in \mathbb{R}$, l'estimateur à noyau de la densité $\rho_\infty f$ en y_0, défini par

$$p_\gamma(y_0) = \frac{\sum_{k\in U_\gamma} I_{\gamma k} K\left(\frac{Y_k - y_0}{h_\gamma}\right)}{h_\gamma n_\gamma},$$

où $n_\gamma = \sum_{k=1}^{N_\gamma} I_{\gamma k}$, vérifie

$$\begin{aligned} \mathrm{E}\left[p_\gamma(y_0) - (\rho_\infty f)(y_0)\right] &= \frac{h_\gamma^2}{2}\left(\left(\frac{\partial^2}{\partial y^2}(\rho_\gamma \times f)\right)(y_0)\int u^2 K(u)du\right) \\ &\quad + o_\gamma(h_\gamma^2) \\ \textit{et } \mathrm{Var}\left[p_\gamma(y_0)\right] &= \left(\left(\frac{v_\infty}{\tau^2} + \rho_\infty^2\right)(y_0)\right)\left(\frac{f(y_0)}{N_\gamma h_\gamma}\int K^2(u)du\right) \\ &\quad + o_\gamma\left(\frac{1}{N_\gamma h_\gamma}\right), \end{aligned}$$

où $\tau = \int(\lim_\gamma m_\gamma) f \mathrm{d}\lambda$

Ce résultat permet le calcul de $\mathrm{Var}\left[p_\gamma(y_0)/\rho_\infty(y_0)\right]$. Ce résultat est précédé d'une généralisation du lemme de Bochner au cas de non indépendance des observations. Il est intéressant de noter que nous retrouvons la vitesse de convergence classique du cas d'un plan de sondage non informatif car indépendant de Y : dans ce cas, $m_\gamma(y)$ ne dépend pas de y,

$$\frac{v_\infty}{\tau^2} + \rho_\infty^2 = \tau^{-1}$$

et

$$\left(\frac{\partial^2}{\partial y^2}(\rho_\gamma \times f)\right)(y_0) = \left(\frac{\partial^2}{\partial y^2} f\right)(y_0).$$

Lorsque ρ_∞ est connnu, $p_\gamma(y)/\rho_\infty(y)$ peut être comparé à l'estimateur de Horvitz-Thompson ou à l'estimateur de type Hájek de $f(y_0)$, defini dans le cas d'un sondage sans remplacement comme

$$\left(\sum_{k\in U_\gamma}\frac{I_{\gamma k}}{\pi_{\gamma k}}\right)^{-1}\left(\frac{1}{h_\gamma}\sum_{k\in U_\gamma}\frac{I_{\gamma k}K\left(\frac{Y_k - y_0}{h_\gamma}\right)}{\pi_{\gamma k}}\right),$$

où $\pi_{\gamma k}$ est la probabilité d'inclusion du k dans l'échantillon sur la γ^{e} population. Celà permet donc de construire des estimateurs de la densité sur la population qui constituent une alternative aux estimateurs qui utilisent des méthodes de substitution qui font intervenir l'estimateur de Horvitz-Thompson.

Dans le chapitre 5, nous considérons une version paramétrique du modèle de superpopulation décrit dans le chapitre 2. Nous supposons que la densité de la variable d'intérêt Y sur la population est f_θ, et que la loi de la variable auxiliaire Z utilisée pour le plan conditionnelle à la variable d'intérêt est indexée par un paramètre ξ. Le but de l'inférence est l'estimation de θ. Nous supposons que $\mathrm{P}_{\theta,\xi}^{Y_k,Z_k}$ admet une densité par rapport à la mesure de Lebesgue. Nous supposons que $\mathrm{P}_{\theta,\xi}^{Y_k}$ ne dépend pas de ξ et nous notons $f_\theta = d\,\mathrm{P}_{\theta,\xi}^{Y_k}/\mathrm{d}\lambda_p$ cette densité. Nous supposons encore que $\mathrm{P}_{\theta,\xi}^{Z_k|Y_k}$ ne dépend pas de θ. Nous définissons, comme au chapitre 2, la loi de distribution de l'échantillon, qui fait dorénavant intervenir une fonction de poids paramêtrée par θ et ξ :

$$\rho_{\gamma,\theta,\xi}(y) = \frac{\mathrm{E}_{\theta,\xi}\left[I_{\gamma k}|Y_k = y\right]}{\mathrm{E}_{\theta,\xi}\left[I_{\gamma k}\right]}.$$

Nous définissons ensuite une approximation de la log-vraisemblance :

$$\begin{aligned}\overline{\mathscr{L}_\gamma}\left(\theta,\xi,\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}}\right)&\\ =\ &\left(n_\gamma+\mathbb{1}_{\{0\}}\left(n_\gamma\right)\right)^{-1}\sum_{k=1}^{n_\gamma}\Delta\left(Y_{R_\gamma(k)},\theta,\xi\right)\\ =\ &\left(n_\gamma+\mathbb{1}_{\{0\}}\left(n_\gamma\right)\right)^{-1}\sum_{k=1}^{N_\gamma}I_{\gamma k}\Delta\left(Y_k,\theta,\xi\right),\end{aligned}$$

avec $\Delta(y,\theta,\xi)=\ln\left(\rho_{\infty,\theta,\xi}(y)f_\theta(y)\right)$, et $Y_{R_\gamma(k)}$ désigne le vecteurs des observations sur l'échantillon, où les valeurs des unités sélectionnées plusieurs fois apparaissent plusieurs fois. Autrement dit, on considère que la vraisemblance de l'échantillon est celle d'un échantillon iid de loi $\rho_{\infty,\theta,\xi}.f$.

Pour ξ donné, l'estimateur du maximum de la vraisemblance de l'échantillon approchée de θ adapté à ξ est défini comme :

$$\hat{\theta}_\gamma(\xi)=\arg\max_{\theta\in\Theta}\left\{\overline{\mathscr{L}_\gamma}\left(\theta,\xi,\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}}\right)\right\}.$$

Nous supposons que nous disposons d'un estimateur consistant $\hat{\xi}$ de ξ. Cet estimateur est alors substitué à ξ dans l'expression de l'approximation de la vraisemblance de l'échantillon et nous étudions les propriétés de l'estimateur $\hat{\theta}\left(\hat{\xi}\right)$ qui maximise cette expression en θ. A partir d'une adaptation de Gong and Samaniego (1981), nous prouvons l'existence, la consistance et le taux de convergence vers une loi normale de cet estimateur :

Théorème 4. *Sous certaines conditions de régularité, et d'indépendance asymptotique de sélection de deux individus (cf théorème 5.1, p.62 et théorème 5.2, p.62), étant donnés θ_0,ξ_0, A (resp. B) un voisinage de θ_0 (resp. ξ_0), $\hat{\xi}_\gamma$ une suite de variables aléatoires telle que $\hat{\xi}_\gamma-\xi_0=o_{\mathrm{P}_{\theta_0,\xi_0}}(1)$, pour $\gamma\in\mathbb{N},\varepsilon\in\mathbb{R}^+$, en notant $C_\gamma(\varepsilon)$ l'évènement*

$$\left\{\exists\hat{\theta}_\gamma\ tel\ que\ \sum_{k=1}^{N_\gamma}\left(\partial\Delta/\partial\theta\right)\left(Y_k,\theta,\hat{\xi}_\gamma\right)I_{\gamma k}=0\ et\ \left|\hat{\theta}_\gamma-\theta_0\right|<\varepsilon\right\},$$

alors

$$\lim_{\gamma\to\infty}\mathrm{P}_{\theta_0,\xi_0}\left(C_\gamma(\varepsilon)\right)=1$$

et si

$$\sqrt{n_\gamma}\begin{bmatrix}\left(\frac{\partial}{\partial\theta}\overline{\mathscr{L}}\right)\left(\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)\\ \hat{\xi}_\gamma-\xi_0\end{bmatrix}\xrightarrow[\gamma\to\infty]{\mathscr{L}}\mathscr{N}\left(0,\Sigma=\begin{bmatrix}\Sigma_{11}&\Sigma_{12}\\ &\Sigma_{22}\end{bmatrix}\right),$$

alors

$$\sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)/\sigma\xrightarrow[\gamma\to\infty]{\mathscr{L}}\mathscr{N}(0,1),$$

où

$$\sigma^2=\frac{\Sigma_{11}}{\mathscr{I}_{11}^2}+\frac{\mathscr{I}_{12}}{\mathscr{I}_{11}^2}\left(\Sigma_{22}\mathscr{I}_{12}-2\Sigma_{12}\right),$$

$$\mathscr{I}_{11}=\int_{\mathbb{R}^p}\left(\frac{\partial\Delta}{\partial\theta}\left(y,\theta_0,\xi_0\right)\right)^2 d\left(\rho_{\infty,\theta_0,\xi_0}f_{\theta_0}.\lambda_p\right)(y),$$

et

$$\mathscr{I}_{12}=\int_{\mathbb{R}^p}\left(\frac{\partial\Delta}{\partial\xi}\frac{\partial\Delta}{\partial\theta}\right)\left(y,\theta_0,\xi_0\right)\ d\left(\rho_{\infty,\theta_0,\xi_0}f_{\theta_0}.\lambda_p\right)(y).$$

Comme dans les chaptires 3 et 4, les conditions que nous proposons sont vérifiables pour une liste de plans de sondage, rencontrés couramment dans la pratique des enquêtes par sondage. À partir d'un jeu d'hypothèses fortes, (notamment l'hypothèse selon laquelle la taille de l'échantillon doit rester fixe et la taille de la population doit tendre vers l'infini), Pfeffermann et al. (1998) ont établi la convergence ponctuelle de la loi jointe des réponses sur l'échantillon vers le produit des densités de l'échantillon. Ce résultat a été avancé comme une justification partielle de la méthode basée sur l'approximation de la vraisemblance de l'échantillon par le produit des densités de l'échantillon. D'autres méthodes de maximisation d'un critère existent, notamment la maximisation de la vraisemblance exacte des données (Breckling et al., 1994; Chambers et al., 1998), ou la maximisation de la *pseudo-vraisemblance*, qui résulte de la substitution de totaux par leurs estimateurs de Horvitz-Thompson dans l'expression de la vraisemblance des données sur toute la population (e.g. Binder, 1983; Chambers and Skinner, 2003; Kish and Frankel, 1974, §2.4). Nous montrons que l'utilisation de la vraisemblance obtenue comme produit de densités sur l'échantillon est possible et permet l'obtention d'etimateurs qui ont une variance moindre que les estimateurs qui maximisent la pseudo-vraisemblance. Nous illustrons ce résultat final avec des simulations appliquées au sondage stratifié, pour lequel nous montrons que les conditions pour la convergence des estimateurs sont vérifiées. Pour celà, nous montrons un théorème de normalité asymptotique pour des sommes sur l'échantillon dans le cas particulier du sondage stratifié avec un nombre fixé (non aléatoire et ne variant pas avec la taille de la population) de strates. Nous calculons dans ce cas les valeurs de σ^2 et comparons la variance ainsi calculée à un calcul de la variance obtenu par la méthode de Monte Carlo à partir de simulations.

Dans le chapitre 6, des résultats sur le sondage équilibré (voir Deville and Tillé (2004, 2005)) sont présentés, sous le modèle fixe de population. Ce travail est le résultat d'une collaboration avec G. Chauvet et J.C. Deville. Ce chapitre se démarque des chapitres précédents, car les résultats exposés ici correspondent à un problème différent du problème principal posé dans la thèse. Les résultats consistent en la présentation de deux algorithmes de calcul de probabilités d'inclusion optimales pour un sondage équilibré.

Deville and Tillé (2005) ont proposé une formule d'approximation de la variance de l'estimateur de Horvitz-Thompson d'un total d'une variable d'intérêt y, (y est un vecteur non aléatoire des valeur d'une caractéristique pour tous les individus de la population) dans le cas d'un tirage équilibré. L'estimateur de Horvitz-Thompson du total est :

$$\hat{t}_y = \sum_{k=1}^{N} y_k \frac{I_k}{\pi_k},$$

sa variance est égale à :

$$\mathrm{Var}\left[\hat{t}_{y\pi}\right] = \sum_{k,l \in U} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} (\pi_{kl} - \pi_k \pi_l).$$

Un plan de sondage p est équilibré sur les variables $\mathbf{x}$ si

$$\mathrm{p} - a.s(i), \sum_{k \in U} \mathbf{x}_k \frac{i_k}{\pi_k} = t_{\mathbf{x}},$$

où $t_{\mathbf{x}}$ désigne le vecteur des totaux sur la population des variables d'équilibrage.

Etant donné un tirage équilibré sur un ensemble de variables auxiliaires $\mathbf{x}$, qui peuvent dépendre des probabilités d'inclusion, la variance approximée est une fonction de y, $\mathbf{x}$, et des probabilités d'inclusion :

$$V_{app}\left(y, \pi, \mathbf{x}\right) = \frac{N}{N-q} \sum_{k \in U} b(\pi_k) \left(y_k - y_k^*(\pi, \mathbf{x})\right)^2,$$

où q désigne le nombre de variables d'équilibrage, $b(\pi_k) = 1/\pi_k - 1$ et $(y_k^*(\pi), \mathbf{x})_{k\in U} = \mathbf{x}_k\ \beta(\pi)$ est la projection orthogonale de y sur l'espace engendré par les variables d'équilibrage, où le produit scalaire est donné par la matrice définie positive $W(\pi)$ dépendant de π, définie par

$$[W(\pi)]_{k,k'} = \begin{vmatrix} (\pi_k)^{-1} - 1 & \text{si } k = k' \\ 0 & \text{sinon.} \end{vmatrix}$$

Autrement dit,

$$V_{app}\left(y, \pi, \mathbf{x}\right) \quad = \quad \left\| \left(\left(W(\pi)\right)^{1/2} \left(\mathrm{Id}_N - \mathbf{x} \left(\mathbf{x}^{\mathrm{T}}\, W(\pi)\, \mathbf{x} \right)^{-1} \mathbf{x}^{\mathrm{T}}\, W(\pi) \right) \right) y \right\|^2 .$$

Lorsque les variables d'équilibrage dépendent elles-même de π, la variance approchée devient :

$$V_{app}\left(y, \pi, \mathbf{x}\left(\pi\right)\right) \quad = \quad \left\| \left(\left(W(\pi)\right)^{1/2} \left(\mathrm{Id}_N - \mathbf{x}\left(\pi\right) \left(\mathbf{x}\left(\pi\right)^{\mathrm{T}}\, W(\pi)\, \mathbf{x}\left(\pi\right) \right)^{-1} \mathbf{x}\left(\pi\right)^{\mathrm{T}}\, W(\pi) \right) \right) y \right\|^2 .$$

Etant donné un jeu de contraintes linéaires $A\pi = a$ sur les probabilités d'inclusions (par exemple, contrainte d'espérance de taille d'échantillon égale à n : $\sum_{k\in U} \pi_k = n$, on appelle optimales les probabilités d'inclusion pour lesquelles cette approximation de la variance est minimale, c'est à dire le vecteur :

$$\pi^* = \arg\min \left\{ V_{app}(y, \pi, \mathbf{x}\left(\pi\right)) \,\middle|\, \pi \in \left]0,1\right]^N \text{ tel que } A\pi = a \right\}.$$

Le premier des algorithmes de calcul est proposé pour le cas particulier où les variables d'équilibrage sont qualitatives et ne dépendent pas des probabilités d'inclusion. L'existence d'une solution globale au problème d'optimisation et la convergence de l'algorithme vers ce minimum sont prouvées.

Le second algorithme est proposé pour le cas où les variables d'équilibrage dépendent des probabilités d'inclusion. La convergence vers un minimum local est prouvée.

Pour le premier algorithme, des simulations sont réalisées, qui montrent que dans certains cas, la minimisation de l'approximation de la variance s'accompagne de la minimisation de la vraie variance, estimée à partir de la méthode de Monte-Carlo. Il est intéressant de noter que les résultats obtenus appliqués au cas ou $\mathbf{x}$ est réduit à une variable qualitative à plusieurs modalités, et où on impose la contrainte de taille d'échantillon on retrouve approximativement le résultat de Neyman, à savoir que les probabilités d'inclusion optimales sont proportionnelles à la racine carrée de la dispersion dans les strates formées selon les différentes modalités de $\mathbf{x}$.

Une partie de ce travail a été publiée dans la revue *Journal of Statistical Planning and Inference* (Chauvet et al., 2011). Une version plus complète a été publiée dans les *Actes des Journées de Méthodologie Statistique of the Insee* (Bonnéry et al., 2009).

Les preuves longues sont données en annexe. L'annexe A présente les concepts et notations mathématiques de base utilisées et non définies dans le reste du document. Un index des notations et un index des termes est fourni à partir de la page 146.

References

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Journal of Statistical Planning and Inference*, 51(3):279–292.

Bonnéry, D., Breidt, F. J., and Coquet, F. (2011). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli*. To appear.

Bonnéry, D., Chauvet, G., and Deville, J.-C. (2009). Optimum de type neyman pour l'échantillonnage équilibré sur des marges. In *Actes des Journées de Méthodologie Statistique*.

Breckling, J., Chambers, R. L., Dorfman, A. H., Tam, S., and Welsh, A. (1994). Maximum likelihood inference from sample survey data. *International Statistical Review/Revue Internationale de Statistique*, 62(3):349–363.

Cassel, C.-M., Särndal, C.-E., and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley-Interscience [John Wiley & Sons], New York. Wiley Series in Probability and Mathematical Statistics.

Chambers, R. L., Dorfman, A. H., and Wang, S. (1998). Limited information likelihood analysis of survey data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(2):397–411.

Chambers, R. L. and Skinner, C. J. (2003). *Analysis of Survey Data*. Wiley Series in Survey Methodology. John Wiley & Sons Inc, Chichester.

Chauvet, G., Bonnéry, D., and Deville, J.-C. (2011). Optimal inclusion probabilities for balanced sampling. *Journal of Statistical Planning and Inference*, 141(2):984–994.

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912.

Deville, J.-C. and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128(2):569–591.

Eideh, A. A. H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *Journal of Statistical Planning and Inference*, 136(9):3052–3069.

Eideh, A. A. H. and Nathan, G. (2007). Corrigendum to "fitting time series models for longitudinal survey data under informative sampling". *Journal of Statistical Planning and Inference*, 137(2):628.

Eideh, A. A. H. and Nathan, G. (2009). Two-stage informative cluster sampling-estimation and prediction with applications for small-area models. *Journal of Statistical Planning and Inference*, 139(9):3088–3101.

Gong, G. and Samaniego, F. J. (1981). Pseudomaximum likelihood estimation: theory and applications. *The Annals of Statistics*, 9(4):861–869.

Kish, L. and Frankel, M. (1974). Inference from complex surveys. *Journal of the Royal Statistical Society, Series B*, 36(1):1–37.

Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8(4):1087–1114.

Pfeffermann, D., Moura, F. A. D. S., and do Nascimento Silva, P. L. (2006). Multi-level modelling under informative sampling. *Biometrika*, 93(4):943–959.

Pfeffermann, D. and Sverchkov, M. Y. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhyā: The Indian Journal of Statistics, Series B*, 61(1):166–186.

Pfeffermann, D. and Sverchkov, M. Y. (2003). Fitting generalized linear models under informative sampling. In *Analysis of survey data*, Wiley Series in Survey Methodology, pages 175–195. Wiley, Chichester.

Pfeffermann, D. and Sverchkov, M. Y. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association*, 102(480):1427–1439.

Pfeffermann, D. and Sverchkov, M. Y. (2009). Inference under Informative Sampling. In Pfefferman, D. and Rao, C., editors, *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of Statistics*, pages 455–487. Elsevier.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer-Verlag, New York.

Skinner, C. J. (1994). Sample models and weights. In statistical association, A., editor, *Proceedings of the Section on Survey Research Methods*, pages 133–142, Washington, DC.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Chapter 2

Informative selection and sample distribution

In this chapter, we define a general statistical model that is suitable for both model-based and design-based inference. Here, model-based inference refers to inference based on the superpopulation model, in which the characteristics of the elements of the population are considered random, and the aim of inference is the random process that generates those characteristics. Design-based inference refers to inference based on the fixed population model, in which the characteristics are considered fixed and constitute the target of the inference, and only the randomness of the sample is used in making inference.

Here we define an exchangeable population model that includes both the superpopulation model and the fixed population model. Under the exchangeable population model, we give a theoretical definition of informative selection that generalizes the definition commonly accepted in model-based inference.

We also give a general definition of the sample distribution under both with and without replacement sampling. We then introduce an asymptotic framework under which we define a limit sample distribution.

2.1 Population, samples, design measures and inclusion probabilities

Let $N \in \mathbb{N}\backslash\{0\}$. A population of size N is a set U of cardinal number N, and for simplicity, a population of size N will always be the set $U = \{1, \ldots, N\}$ in the following. A sample i from the population U is a vector in $\mathbb{N}^N$. It is standard in the survey literature to consider a sample as a subset of U. For simplicity of notation, we define a sample i from the population U as an element of $\mathbb{N}^N$. The kth coordinate of the sample i, i_k, indicates the number of times the element k is selected: for example, the vector $i = (3, 0, 5)$ is a sample from the population $U = \{1, 2, 3\}$, in which the element labelled 1 has been selected 3 times, the element labelled 2 has not been selected and the element labelled 3 has been selected 5 times.

In the literature, a design measure, is a function p mapping any subset of U to $[0, 1]$. In this dissertation, p will instead denote a probability measure p on the measurable space $\left(\mathbb{N}^N, \mathscr{P}\left(\mathbb{N}^N\right)\right)$, where $\mathscr{P}\left(\mathbb{N}^N\right)$ is the power set of $\mathbb{N}^N$. This measure will be called the design measure. The set of design measures on U is denoted $\mathbb{\Pi}$.

Example 2.1. *Simple random sampling*
Let $n \in \{0, \ldots, N\}$. Simple random sampling without replacement of n elements from N elements is characterized by:

$$\forall i \in \mathbb{N}^N,\ \mathrm{SRS}_{N,n}(\{i\}) = \begin{cases} \left(\binom{N}{n}\right)^{-1} & \text{if } \sum_{k\in U} i_k = n \text{ and } i \in \{0,1\}^N, \\ 0 & \text{otherwise.} \end{cases}$$

Example 2.2. *Poisson sampling*
Let $z \in [0,1]^N$. The Poisson sampling with inclusion probabilities z is the design measure Pois_z characterized by:

$$\forall i \in \mathbb{N}^N, \mathrm{Pois}_z\left(\{i\}\right) = \begin{cases} \prod_{k=1}^N z_k^{i_k}(1-z_k)^{1-i_k} & \text{if } i \in \{0,1\}^N, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

A sample from the population U drawn according to the design measure p is a random vector $\mathcal{I} = (I_k)_{k\in U}$ with value in $\mathbb{N}^N$ such that $\mathcal{I} \sim \mathrm{p}$.

We distinguish *with replacement* and *without replacement* design measures :

Definition 2.1. *For $\mathrm{p} \in \mathbb{\Pi}$, p is a design measure without replacement if $\mathrm{p}\left(\{0,1\}^N\right) = 1$, otherwise it is with replacement.*

For example, simple random sampling and Poisson sampling are design measures without replacement.

2.2 Study variables and design variables

In the following, if not otherwise specified, all random variables are defined on the same probability space $(\Omega, \mathscr{A}, \mathrm{P})$.

We define two vectors of random variables: $\mathcal{Y} = (Y_k)_{k\in U}$ and $\mathcal{Z} = (Z_k)_{k\in U}$, where for $k \in U$, $Y_k : (\Omega, \mathscr{A}) \to (\mathscr{Y}, \mathscr{T}_Y)$ and $Z_k : (\Omega, \mathscr{A}) \to (\mathscr{Z}, \mathscr{T}_Z)$, with $(\mathscr{Y}, \mathscr{T}_Y)$ and $(\mathscr{Z}, \mathscr{T}_Z)$ being measurable spaces. For the kth element of U, Y_k corresponds to the characteristics of interest, and Z_k corresponds to the design characteristics. We denote $\mathcal{Y} = (Y_k)_{k\in U}$ and $\mathcal{Z} = (Z_k)_{k\in U}$.

Let Π be a random design measure on the population U, such that Π is a function of the design variable. Such a description of the population model can be found in Skinner (1994, pp. 134-135). Assume that the index of the element k of the population plays no role in the way elements are selected. Specifically:

$$\begin{cases} \Pi : \Omega \to \mathbb{\Pi} & \text{(2.2a)} \\ \exists s \in \mathscr{F}\left(\mathscr{Z}^N, \mathbb{\Pi}\right) \text{ such that } \Pi = D(\mathcal{Z}) & \text{(2.2b)} \\ \forall z \in \mathscr{Z}^N,\ r \text{ a permutation of } U,\ A \subset \mathbb{N}^N,\ (D(z))(A) = (D(r.z))(r.A), & \text{(2.2c)} \end{cases}$$

where $r.z = \left(z_{r(1)} \dots z_{r(N)}\right)$, $r.A = \left\{\left(a_{r(1)} \dots a_{r(N)}\right) \middle| a \in A\right\}$, and $\mathscr{F}\left(\mathscr{Z}^N, \mathbb{\Pi}\right)$ is the set of functions from $\mathscr{Z}^N$ to $\mathbb{\Pi}$. In this notation, $\Pi = D(\mathcal{Z})$ is a random design measure, $D(z)$ is a realization of the random design measure, and D is called the design measure function. Equation (2.2c) implies that $\forall r$ a permutation of U, A a subset of $\mathbb{N}^N$, $\omega_1, \omega_2 \in \Omega$:

$$\mathcal{Z}(\omega_2) = r.\mathcal{Z}(\omega_1) \Rightarrow \Pi(\omega_1)(A) = (\Pi(\omega_2))(r.A). \tag{2.3}$$

That is, if the design information z were randomly reordered, then the measure of the reordered samples in A would be invariant.

Example 2.3. *Poisson sampling*
Assume $\mathscr{Z} = [0,1]$. Then by equation (2.1), $\mathrm{Pois}_{\mathcal{Z}}$, the Poisson sampling with inclusion probabilites $\mathcal{Z}$, verifies (2.2).

The random vector $\mathcal{I}$ continues to take values in $\mathbb{N}^N$, but now denotes a sample from U selected according to the random design measure Π, with:

$$\begin{cases} \mathrm{P}^{\Pi,\mathcal{Y},\mathcal{Z}} - a.s.(\mathrm{p}, y, z), \mathrm{P}^{\mathcal{I}|\Pi=\mathrm{p},\mathcal{Y}=y,\mathcal{Z}=z} = \mathrm{P}^{\mathcal{I}|\Pi=\mathrm{p}} & (2.4a) \\ \mathrm{P}^{\Pi} - a.s.(\mathrm{p}),\ \mathrm{P}^{\mathcal{I}|\Pi=\mathrm{p}} = \mathrm{p}, & (2.4b) \end{cases}$$

where $\mathrm{P}^{\mathcal{I}|\Pi=\mathrm{p}} = \mathrm{p}$ is the distribution of $\mathcal{I}$ conditional on $\Pi = \mathrm{p}$. Equation (2.4b) means that $\forall A \in \mathscr{P}\left(\mathbb{N}^N\right)$ and $\forall B$ in the smallest σ-algebra on $\mathbb{\Pi}$ that contains $\Pi(\mathscr{A})$,

$$\mathrm{P}^{I,\Pi}(A \times B) = \int_B \mathrm{p}(A)\, d\,\mathrm{P}^{\Pi}(\mathrm{p}).$$

Equation (2.4) expresses the idea that those responsible for the sampling procedure have defined a design measure as a function of the design variables at their disposal. Once the design measure was defined, the sample was drawn without any further use of the design variables or study variables.

Definition 2.2. *Inclusion and double inclusion probabilities*
We define the inclusion probability of element $k \in U$ as the random variable:

$$\pi_k = \Pi\left(\left\{ i \in \mathbb{N}^N \middle| i_k \geqslant 1 \right\}\right),$$

and the second order inclusion probability of elements k and l as the random variable:

$$\pi_{k,l} = \Pi\left(\left\{ i \in \mathbb{N}^N \middle| i_k \geqslant 1, i_l \geqslant 1 \right\}\right).$$

Define the sample size as the random variable $n = \sum_{k=1}^{N} I_k$. We now define a random variable R that allows distinction between labelled and unlabelled samples:

$$\begin{cases} \forall \omega \in \Omega,\ R(\omega) \in \{r : \{1,\ldots,n\} \to U \mid \forall k \in U, \#\left(\{l \in \{1,\ldots,n(\omega)\} \mid r(l) = k\}\right) = I_k(\omega)\} & (2.5a) \\ \mathrm{P}^{R|\mathcal{I}=i,\mathcal{Y}=y,\mathcal{Z}=z} = \mathrm{P}^{R|\mathcal{I}=i}\, \mathrm{P}^{\mathcal{I},\mathcal{Y},\mathcal{Z}} - a.s(i, y, z) & (2.5b) \\ \mathrm{P}^{R|\mathcal{I}=i} \text{ is the uniform law on} \\ \qquad \{r : \{1,\ldots,n\} \to U | \forall k \in U, \#(\{l \in \{1,\ldots,n\} | r(l) = k\}) = i_k\}. & (2.5c) \end{cases}$$

To a sample $\mathcal{I}$ can be associated a list of observations that consists of a file of n lines, each line corresponding to one sampled element of the population. When the sampling is with replacement, some lines may be replicated. This file has its own labels, which are numbers from 1 to n. The equation $k = R(\ell)$ means that the ℓth observation (the ℓth line) corresponds to the kth element of the population ; equation (2.5a) means that the number of observations ℓ that correspond to the kth element of the population equals I_k ; equations (2.5b) and (2.5c) mean that the order in which observations appear is totally random and does not depend either on the random design measure, the design variables, or the study variables, given the number of times each element is selected. The vector $R.\mathcal{Y} = \left(Y_{R(1)} \ldots Y_{R(n)}\right)$ corresponds to the vector of the unlabelled study variables for the selected elements: if the analyst just observes $R.\mathcal{Y}$, and does not know Π, $\mathcal{I}$ or R, then the analyst does not know to which elements in the population the observations $\left(Y_{R(\ell)}\right)_{\ell \in \{1,\ldots,n\}}$ correspond.

We can assume that $(\Omega, \mathscr{A}, \mathrm{P})$ is of the form

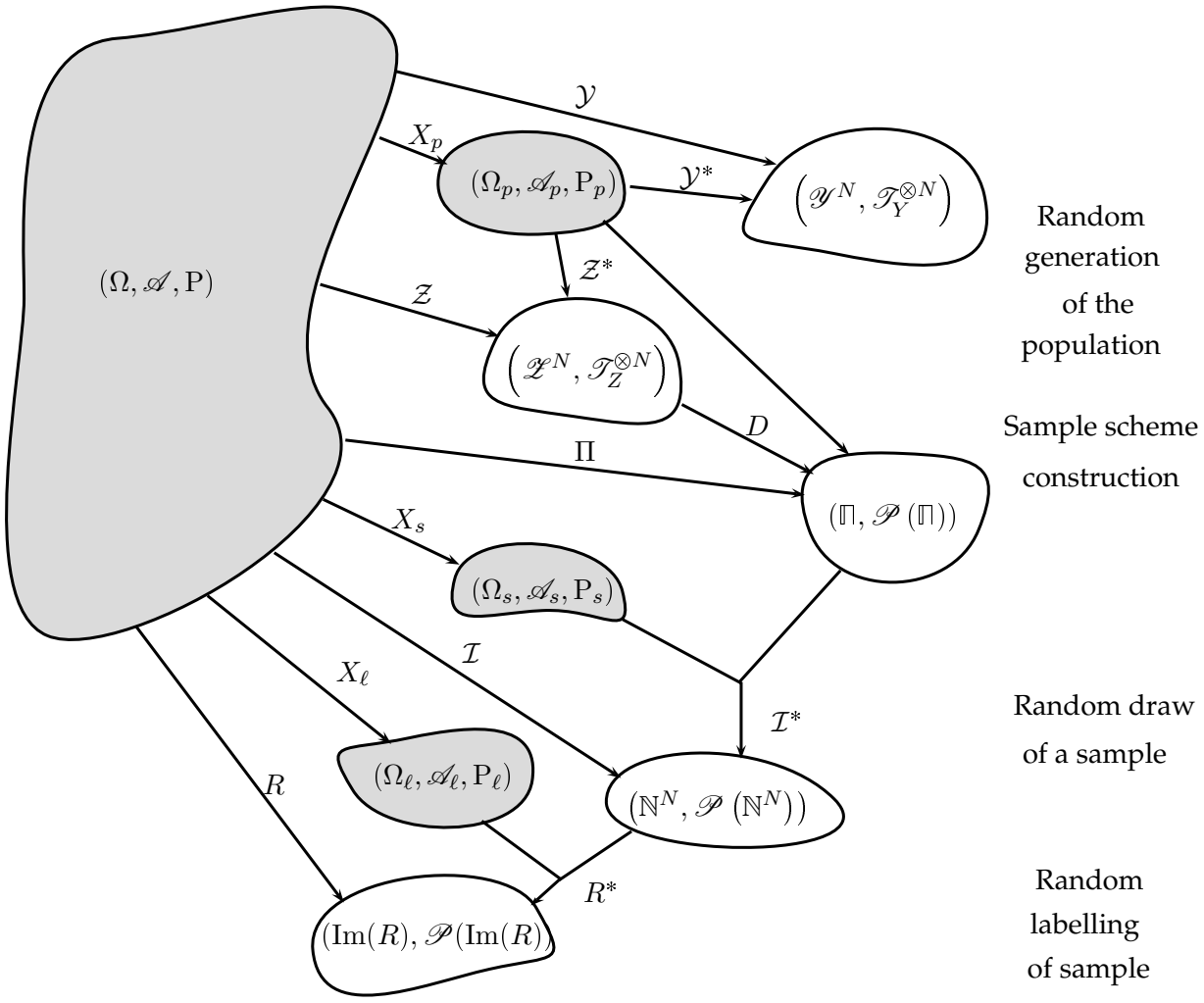
$$(\Omega, \mathscr{A}, \mathrm{P}) = (\Omega_p \times \Omega_s \times \Omega_\ell, \mathscr{A}_p \otimes \mathscr{A}_s \otimes \mathscr{A}_\ell, \mathrm{P}_p \otimes \mathrm{P}_s \otimes \mathrm{P}_\ell),$$

where $(\Omega_p, \mathscr{A}_p, \mathrm{P}_p)$, $(\Omega_s, \mathscr{A}_s, \mathrm{P}_s)$ and $(\Omega_\ell, \mathscr{A}_\ell, \mathrm{P}_\ell)$ are probability spaces. The elements of Ω are of the form $\omega = (\omega_p, \omega_s, \omega_\ell)$, where $\omega_p \in \Omega_p, \omega_s \in \Omega_s, \omega_\ell \in \Omega_\ell$. Let X_p, X_s, X_ℓ denote the projections :$X_p : \omega \mapsto \omega_p$,

$X_s : \omega \mapsto \omega_s$, $X_\ell : \omega \mapsto \omega_\ell$. Assume that $\forall \omega_p \in \Omega_p, \omega'_s, \omega_s \in \Omega_s, \omega_\ell, \omega_\ell, \in \Omega_\ell$, $\mathcal{Y}(\omega_p, \omega_s, \omega_\ell) = \mathcal{Y}(\omega_p, \omega'_s, \omega'_\ell)$, $\mathcal{Z}(\omega_p, \omega_s, \omega_\ell) = \mathcal{Z}(\omega_p, \omega'_s, \omega'_\ell)$. We can then define $\mathcal{Y}^*$, $\mathcal{Z}^*$ such that $\mathcal{Y} = \mathcal{Y}^* \circ X_p$, $\mathcal{Z} = \mathcal{Z}^* \circ X_p$.

Next, we assume that there exists a measurable function $\mathcal{I}^* : (\mathbb{\Pi} \times \Omega_s, \mathscr{P}(\mathbb{\Pi}) \otimes \mathscr{A}_s) \to \left(\mathbb{N}^N, \mathscr{P}\left(\mathbb{N}^N\right)\right)$, $(\mathrm{p}, \omega_s) \mapsto \mathcal{I}^*(\mathrm{p}, \omega_s)$ such that $\mathcal{I}(\omega_p, \omega_s, \omega_\ell) = \mathcal{I}^*(\Pi(\omega_p, \omega_s, \omega_\ell), \omega_s)$. Then, $\forall \mathrm{p} \in \mathbb{\Pi}, \forall A \in \mathscr{P}\left(\mathbb{N}^N\right)$, $\mathrm{p}(A) = \mathrm{P}_s\left(\{\omega_s \in \Omega_p | \mathcal{I}^*(\mathrm{p}, \omega_s) \in A\}\right)$. Define . We assume that $(\Omega_s, \mathscr{A}_s, \mathrm{P}_s) = \left([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]}\right)$. We assume that there exists R^* such that $R = R^*(X_\ell, \mathcal{I})$. The commutative diagram of figure 2.1 summarizes what precedes.

Figure 2.1: Commutative diagram for $\mathcal{Y}, \mathcal{Z}, \Pi, \mathcal{I}, R$



What follows focuses on the concept of observation in survey sampling, to get finally to the definition of the general model for survey sampling.

Definition 2.3. *Observation*

The observation is a random variable G that is a function, denoted g, of $\mathcal{I}, \mathcal{Y}, R$:

$$G = g(\mathcal{I}, \mathcal{Y}, R).$$

Many types of observations are possible, in particular the following ones:

- In the case where the analyst just observes a list of responses to a survey, without knowing to which element in the population each response on the list corresponds, or how many times a response is replicated in the sample (the observed data are unlabelled), then $G = R.Y$.
- If, in addition, the analyst knows to which element the response corresponds and how many times each element has been selected, then $G = (Y_k, I_k)_{k \in U | I_k > 0}$ (the observed data are labelled).

Notice that the design variables and the study variables do not need to be separated: a variable can be both a design variable and a study variable.

The probability law of the observations is then P^G, and depending on the aim of inference, a statistical model can be specified for P^G. Due to (2.2) and (2.4), P^G can be deduced from $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}$, g, and D. In all the following examples, for a complete description of the model, we will just need to specify the population model $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}$, the design measure function D, and the observation function g.

2.3 Population model

2.3.1 Parametric fixed population model for design-based inference

In design-based inference for sampling from a finite population, $(\Omega, \mathscr{A}, \mathrm{P})$ belongs to the following parametric model:

$$(\Omega, \mathscr{A}, \mathrm{P}_{y,z})_{(y,z) \in \Theta},$$

where Θ is a subset of $\mathscr{Y}^N \times \mathscr{Z}^N$. We now need to specify $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}$. To keep track of the parameters, we subscript $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}$ and write $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}_{y,z}$. The study variables and the design variables are not random: $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}_{y,z} = \delta_{\{(y,z)\}}$, where $\delta_{\{(y,z)\}}$ is the Dirac measure in (y, z). The design measure is then not random: $\forall \omega \in \Omega$, $\Pi(\omega) = D\left(\mathcal{Z}(\omega)\right) = D(z)$. This model is called the *fixed population model*.

We can distinguish whether the observed data are labelled or unlabelled:

- if $G = \left(\mathcal{I}, (Y_k)_{k \in U | I_k > 0}\right)$ (we know who has been selected, and how many times), then we define the function $g_y : \mathbb{N}^N \to \mathbb{N}^N \times \bigcup_{d \in \{0,\ldots,N\}} \mathbb{R}^d$, $i \mapsto \left(i, (y_k)_{k \in U | i_k > 0}\right)$,
- if $G = \left(\left(Y_{R(\ell)}\right)_{\ell \in \{1,\ldots,n\}}\right)$ (we do not know neither who has been selected, nor how many times each element in the sample has been selected), then we define g_y as the function $g_y : \bigcup_{d \in \mathbb{N}} \mathscr{F}\left(\{1, \ldots, d\}, U\right) \to \bigcup_{d \in \mathbb{N}} \mathbb{R}^d, r \mapsto r.y$.

Then (see Gourieroux (1981, p. 52)) the model used for inference on y is:

$$\mathrm{P}^G_{y,z} = (D(z))^{g_y}. \tag{2.6}$$

This model is called the *design-based model*. In this parametric model, "the only stochastic element upon which an inference can be made is the one introduced through the [design measure]" (Cassel et al., 1977, p. 32). Usually z is known, the study characateristics are the parameters of the model, and the target of the inference is a function of the parameter y.

Example 2.4. *Simple random sampling of size n^**
In this example, n is not random, and takes the value n^*, with $n^* \in \mathbb{N} \backslash \{0\}$. The observation is $G = R.\mathcal{Y}$,

and the design measure Π *is non-random:* $\forall z, D(z) = \mathrm{SRS}_{N,n^*}$. *Then the parametric model to consider is:*

$$\mathrm{P}_y^G = \frac{\sum_{r \in \mathscr{I}(\{1,\ldots,n^*\},U)} \delta_{r.y}}{\binom{N}{n^*}},$$

where $\mathscr{I}(\{1,\ldots,n^*\},U)$ *is the set of injective functions from* $\{1,\ldots,n^*\}$ *to* U.

Under the parametric model for design-based inference, a basic tool is the Horvitz-Thompson estimator of the finite population total (Horvitz and Thompson, 1952).

Definition 2.4. *Horvitz Thompson estimator of* $\sum_{k=1}^N y_k$
If from G *the analyst can extract* $(I_k, \pi_k, Y_k)_{k \in U, I_k \geqslant 1}$, *if* $\mathscr{Y}$ *is a* $\mathbb{R}$*-vector space, if* $\forall k \in U$, $\pi_k > 0$, *then the totals of the study variables on the population can be unbiasedly estimated by the Horvitz-Thompson estimator:* $\sum_{k=1}^N \frac{Y_k \mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)}{\pi_k}$.

2.3.2 The iid superpopulation model for model-based inference

An important superpopulation model in model-based inference is one in which the random variables $(Y_1, Z_1), \ldots, (Y_N, Z_N)$ are independent and identically distributed, which is denoted:

$$\mathrm{P}^{\mathcal{Y},\mathcal{Z}} = \bigotimes_{k=1}^N \mathrm{P}^{Y_k,Z_k} = \left(\mathrm{P}^{Y_1,Z_1}\right)^{\otimes N}, \tag{2.7}$$

where $\otimes$ is the symbol for tensor product of measures. In model-based inference, specifying a model consists of assuming that P^{Y_1,Z_1} belongs to a specific family of probability laws. Then the target of the inference is P^{Y_1}.

Example 2.5. *Simple random sampling of size* n^*
In this example, $\mathrm{P}^{\mathcal{Y},\mathcal{Z}}$ *follows the iid superpopulation model, the observation is* $G = R.\mathcal{Y}$, *and the design measure* Π *is non-random:* $\forall z, D(z) = \mathrm{SRS}_{N,n^*}$, *for* $n^* \in \{1,\ldots,N\}$. *Then:*

$$\mathrm{P}^G = \left(\mathrm{P}^{Y_1}\right)^{\otimes n^*}.$$

Example 2.6. *Sampling without replacement, independent from* $\mathcal{Y}$
In this example, consider the case where $\mathcal{Y}$ *and* $\mathcal{Z}$ *are independent random variables, and* $G = R.\mathcal{Y}$, Π *is without replacement. Then (see also (Fuller, 2009, Thm. 1.3.1)):*

$$\mathrm{P}^n - a.s.(n^*),\ \mathrm{P}^{G|n=n^*} = \left(\mathrm{P}^{Y_1}\right)^{\otimes n^*}.$$

Example 2.7. *Hájek estimator of* $\mathrm{E}[Y_1]$
If from G *the analyst can extract* $(I_k, \pi_k, Y_k)_{k \in U, I_k \geqslant 1}$, *if* $\mathscr{Y}$ *is a* $\mathbb{R}$*-vector space, if* $\forall k \in U$, $\pi_k > 0$, *then the expected value of* Y_1 *can be estimated by:* $\left(\sum_{k=1}^N \frac{\mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)}{\pi_k}\right)^{-1} \left(\sum_{k=1}^N \frac{Y_k \mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)}{\pi_k}\right)$.

2.3.3 General case: exchangeable population model

We define an exchangeable population model as a model where the random variables $(Y_1, Z_1), \ldots, (Y_k, Z_k)$ are exchangeable: $\forall r \in \{1,\ldots,N\}$, $k_1 \ldots k_r \in \{1,\ldots,N\}$ distinct, $l_1 \ldots l_r \in \{1,\ldots,N\}$ distinct,

$$\mathrm{P}^{\left((Y_{k_1} Z_{k_1}) \ldots (Y_{k_r} Z_{k_r})\right)} = \mathrm{P}^{\left((Y_{l_1} Z_{l_1}) \ldots (Y_{l_r} Z_{l_r})\right)}. \tag{2.8}$$

Equations (2.4) and (2.8) induce exchangeability of the sequence: $((Y_1Z_1, I_1), \ldots, (Y_NZ_N, I_N))$ and will be refered to as the exchangeable assumption. Specifically, $\forall r \in \{1, \ldots, N\}$, $k_1 \ldots k_r \in \{1, \ldots, N\}$ distinct, $l_1 \ldots l_r \in \{1, \ldots, N\}$ distinct,

$$\mathrm{P}^{\left((Y_{k_1}Z_{k_1}, I_{k_1})\ldots(Y_{k_r}Z_{k_r}, I_{k_r})\right)} = \mathrm{P}^{\left((Y_{l_1}Z_{l_1}, I_{l_1})\ldots(Y_{l_r}Z_{l_r}, I_{l_r})\right)}. \tag{2.9}$$

The iid superpopulation model is obviously a particular case of the exchangeable population model. The fixed population model is not a particular case of the exchangeable population model (except when the parameters y and z are constant vectors). We can propose an alternative to the fixed population model for the design-based model to correspond to a population model that satisfies the exchangeability assumption. This will allow us to extend the definition of the weighted density and the definition of non-informative selection to the design-based case. We define the exchangeable fixed population model as: $(\Omega, \mathscr{A}, \mathrm{P}_{y,z})_{(y,z)\in\Theta}$ where $\Theta \subset \mathscr{Y} \times \mathscr{Z}$ and

$$\mathrm{P}_{y,z}^{\mathcal{Y},\mathcal{Z}} = \frac{1}{N!} \sum_{r\in\mathfrak{S}_N} \delta_{(r.y,r.z)}.$$

Consider now the fixed population model: $\left(\Omega, \mathscr{A}, \mathrm{Q}_{y,z}\right)_{y\in\mathscr{Y}^N, z\in\mathscr{Z}^N}$, where

$$\mathrm{Q}_{y,z}^{\mathcal{Y},\mathcal{Z}} = \delta_{(y,z)}$$

Then we have the following property:

Property 2.1.

$$\forall (x, y) \in \mathscr{Y} \times \mathscr{Z},\ \mathrm{Q}_{y,z}^{R.\mathcal{Y}} = \mathrm{P}_{y,z}^{R.\mathcal{Y}}.$$

When $G = R.\mathcal{Y}$ (the observations are unlabelled), the exchangeable fixed population model and the fixed population model are equivalent.

To any population model can be associated an equivalent exchangeable population model. Consider the statistical model: $(\Omega, \mathscr{A}, \mathrm{Q})_{\mathrm{Q}\in\mathcal{Q}}$, The associated population model is: $\left(\mathscr{Y} \times \mathscr{Z}, \mathscr{T}_Y \otimes \mathscr{T}_Z, \mathrm{Q}^{\mathcal{Y},\mathcal{Z}}\right)_{\mathrm{Q}\in\mathcal{Q}}$, To $(\Omega, \mathscr{A}, \mathrm{Q})_{\mathrm{Q}}$, we associate the exchangeable model $(\Omega, \mathscr{A}, \mathrm{P})_{\mathrm{P}\in\mathcal{P}}$, such that

$$\left\{\mathrm{P}^{\mathcal{Y},\mathcal{Z}} \middle| \mathrm{P} \in \mathcal{P}\right\} = \left\{\frac{1}{N!} \sum_{r\in\mathfrak{S}_N} \mathrm{Q}^{r.\mathcal{Y},r.\mathcal{Z}} \middle| \mathrm{Q} \in \mathcal{Q}\right\}.$$

Those models are equivalent in the sense that:

$$\left[\mathrm{P}^{\mathcal{Y},\mathcal{Z}} = \frac{1}{N!} \sum_{r\in\mathfrak{S}_N} \mathrm{Q}^{r.\mathcal{Y},r.\mathcal{Z}}\right] \Rightarrow \left[\mathrm{P}^{R.\mathcal{Y}} = \mathrm{Q}^{R.\mathcal{Y}}\right].$$

So to the population model $\left(\mathscr{Y} \times \mathscr{Z}, \mathscr{T}_Y \otimes \mathscr{T}_Z, \mathrm{Q}^{\mathcal{Y},\mathcal{Z}}\right)_{\mathrm{Q}\in\mathcal{Q}}$, we can associate the exchangeable population model $\left(\mathscr{Y} \times \mathscr{Z}, \mathscr{T}_Y \otimes \mathscr{T}_Z, \mathrm{P}^{\mathcal{Y},\mathcal{Z}}\right)_{\mathrm{P}\in\mathcal{P}}$.

2.4 Informative selection

Informative selection has been defined under both the model (Pfeffermann and Sverchkov, 2009) and the design (Cassel et al., 1977) approaches. It is possible to extend the definition used in the model-based case to the general case, but we show that this extension does not match the common definition of informative selection under the fixed population model.

2.4.1 Informative selection in the general case

We propose a definition of informative selection that applies for a general case, that includes all exchangeable population models, with and without replacement sampling, all observation functions, and that generalizes the existing definitions stated for the iid superpopulation model.

Definition 2.5. *Informative and non-informative selection*
Consider a general exchangeable model $(\Omega, \mathscr{A}, \mathrm{P})_{\mathrm{P}\in\mathcal{P}}$ *as described by* (2.8)*. Consider some observation function* g*,* $G = g(\mathcal{I}, \mathcal{Y}, R)$*. We will say that the selection is non-informative (on* G*) if*

$$\forall\, \mathrm{P} \in \mathcal{P},\ \mathrm{P}^{n}\ -a.s.(n^*),\ \mathrm{P}^{G|n=n^*} = \mathrm{P}^{g\left(\mathcal{I}^*(\mathrm{SRS}_{N,n^*},X_s),\mathcal{Y},R^*\left(X_\ell,\mathcal{I}^*(\mathrm{SRS}_{N,n^*},X_s)\right)\right)}. \tag{2.10}$$

The selection is informative if it is not non-informative on G*.*

Remark In general, the law of G conditioned on a fixed sample size is:

$$\mathrm{P}^{G|n=n^*} = \mathrm{P}^{g(\mathcal{I}^*(D(\mathcal{Z}),X_s),\mathcal{Y},R^*(X_\ell,\mathcal{I}^*(D(\mathcal{Z}),X_s)))|n=n^*}. \tag{2.11}$$

Then the selection is non-informative if the law of G, given $n = n^*$, is the law of G that would have been obtained with a simple random sampling of n^* elements from N instead of via $D(\mathcal{Z})$.

Property 2.2. *If* $G = R.\mathcal{Y}$*, then according to definition 2.5, if the selection is without replacement and if*

$$\forall n^* \in \{1,\dots,N\} \text{ such that } \mathrm{P}\left(n = n^*\right) > 0,\ \mathrm{P}^{R.Y|n=n^*} = \mathrm{P}^{Y_1\dots Y_{n^*}},$$

then the selection is non-informative.

Definition 2.6. *Non-informative selection conditional on some random variable*
Consider the general exchangeable model. Consider some observation function g*. Let* C *be a random variable. We will say that the selection is non-informative conditional on* C *if*

$$\mathrm{P}^{n,C}\ -a.s.(n^*,c),\ \mathrm{P}^{G|n=n^*,C=c} = \mathrm{P}^{g\left(\mathcal{I}^*(\mathrm{SRS}_{N,n^*},X_s),\mathcal{Y},R^*\left(X_\ell,\mathcal{I}^*(\mathrm{SRS}_{N,n^*},X_s)\right)\right)|C=c}. \tag{2.12}$$

Property 2.3.

The random variables $\mathcal{Z}$ *and* $\mathcal{Y}$ *are independent and the selection is without replacement,*
$\Rightarrow$ Π *and* $\mathcal{Y}$ *are independent and the selection is without replacement,*
$\Rightarrow$ $\mathcal{I}$ *and* $\mathcal{Y}$ *are independent and the selection is without replacement,*
$\Rightarrow$ *the selection is non-informative.*

Simple random sampling is always non-informative under the exchangeable population model and a selection is non-informative if the law of Y in the sample, given that the sample size equals n^*, equals the law of Y in the sample that would be obtained if the design measure was a simple random sample of n^* elements from N elements.

2.4.2 Informative selection under the iid superpopulation model

Definitions 2.5 and 2.6 generalize the standard definition of non-informative selection under an iid superpopulation model (see for example (Pfeffermann and Sverchkov, 2009, p. 455)).

Property 2.4. *Under the iid superpopulation model, if $G = R.Y$, the selection is non-informative if*

$$\mathrm{P}^n - a.s.(n^*),\ \mathrm{P}^{R.Y|n=n^*} = \left(\mathrm{P}^{Y_1}\right)^{\otimes n^*}.$$

When the target of the inference is not P^{Y_1} but $P^{Y_1|\varphi(Y_1)}$, when φ is a function on $\mathscr{Y}$, definition 2.6 allows to determine whether the selection is informative. Consider the following example.

Example 2.8. *Let $\mathscr{Y} = \mathbb{R}^2$, and let $Y_k = (Y_{k,1}, Y_{k,2})$. Let $Y_{.1} = (Y_{k1})_{k\in U}$, $Y_{.2} = (Y_{k2})_{k\in U}$, $\varphi : Y_k \mapsto Y_{k,1}$. Assume $\mathrm{P}^{Y_{.1}} = \mathscr{N}(0, \mathrm{Id}_N)$ and $\mathrm{P}^{Y_{.2}|Y_{.1}=y} = \mathscr{N}(y, \sigma^2 \mathrm{Id}_N)$, where Id_N is the $N \times N$ identity matrix. Assume $\mathscr{Z} = Y_{.1}$, and the selection is such that: $I_k = 1$ if $Z_k > 0$, 0 otherwise. In this example, the selection is informative on $R.\mathcal{Y}$ because it has to be taken into account when making inference on P^{Y_1}. But it can be ignored when trying to estimate σ by standard methods because:*

$$\mathrm{P}^{R.Y_{.2}|n=n^*, R.Y_{.1}=y} = \mathscr{N}(y, \sigma^2 \mathrm{Id}_{n^*}). \tag{2.13}$$

According to definition 2.6, equation (2.13) means that conditional on $C = R.Y_{.1}$, the sample is non-informative on $g(\mathcal{I}, \mathcal{Y}, R) = R.Y_{.2}$.

2.4.3 Informative selection under the fixed population model

Property 2.5. *Informative selection under the fixed population model*
Consider the exchangeable fixed population model. Assume $G = R.\mathcal{Y}$. According to definition 2.5, the selection is non-informative if the selection is without replacement, and if

$$\begin{aligned}\forall (y,z) \in \Theta,\ (D(z))^{g_y} &= \sum_{n^*=0}^{N} \mathrm{P}\left(n = n^*\right) . \left(\mathrm{SRS}_{N,n^*}\right)^{g_y} \\ &= \sum_{n^*=0}^{N} \left((D(z))\left(\left\{i \mid \sum i_k = n^*\right\}\right)\right) . \left(\mathrm{SRS}_{N,n^*}\right)^{g_y}.\end{aligned}$$

Example 2.9. *Consider the case where $\mathscr{Y} = \mathbb{R}$, $\mathscr{Z} =]0,1]$, and $D(z) = \mathrm{Pois}_z$. If $\Theta = \mathscr{Y}^N \times \{(\alpha \dots \alpha) \mid \alpha \in [0,1]\}$, then the selection is non-informative. This design is known as Bernoulli sampling. If $\Theta = \mathscr{Y}^N \times \mathscr{Z}^N$, then for all $z \in \mathscr{Z}^N \backslash \{(\alpha \dots \alpha) \mid \alpha \in [0,1]\}$, $y \in \mathscr{Y}^N \backslash \{(\beta \dots \beta) \mid \beta \in \mathscr{Y}\}$, the selection is informative.*

Remark Definition 2.5 of non-informative selection applied to the fixed population model is not consistent with the definition proposed in Cassel et al. (1977, p. 12), which reads, after replacing their p by q for consistency with our notation: *"A sampling design [...] $\mathrm{q}(.)$ is called a non-informative design if and only if $\mathrm{q}(.)$ is a function that does not depend on the y-values associated with the labels in [the sample] s[...]."* This definition is ambiguous. The ambiguity comes from the fact that in Cassel et al. (1977), the model for inference is not complete because the set of parameters is not given, and what is really meant by "does not depend on" is not explained.

In Cassel et al. (1977), the design measure p is fixed, but may depend (non stochastically) on a function of the parameter y. So we consider that in Cassel et al. (1977), the statistical model which is implicitly referred to is parametrized by the couple (p, y) that belongs to some subset Θ' of $\mathbb{\Pi} \times \mathscr{Y}^N$. The statement that p does not depend on $(y_k)_{k\in U, I_k \geqslant 1}$ is not ambiguous in some particular cases:

- In the case where $\Theta' = \{\mathrm{p}_0\} \times B$ where $\mathrm{p}_0 \in \mathbb{\Pi}$, $B \subset \mathscr{Y}^N$, then $i \mapsto \mathrm{p}(\{i\})$ is a constant function of $(y_k)_{k \in U, i_k = 1}$.
- In the case where $\Theta' = \left\{ (\mathrm{Pois}_y, y) \middle| \, y \in \mathscr{Y}^N \right\}$, $\mathscr{Y} = [0,1]$ then the definition indicates that the sample is informative, as the probability to draw a sample depends in part on the values of y on the sample.

But the definition can be ambiguous in cases where the dependence on y is not direct, but is indirectly imposed by some non trivial correspondence between p and y. For example, in the case where $\Theta' = \left\{ (\mathrm{Pois}_z, y) \middle| \, y \in [0,1]^N, z \in [0,1]^N, \|y - z\|^2 \leqslant \frac{1}{2} \right\}$, it is very difficult to apply the Cassel et al. (1977) definition to determine whether the sample design is informative or not.

Nevertheless, according to a certain interpretation of "does not depend on", the Cassel et al. (1977) definition can be understood as:

$$\text{The selection is non-informative if } \exists A \subset \mathbb{\Pi},\ B \subset \mathscr{Y}^N, \text{such that } \Theta' = A \times B.$$

2.5 Asymptotic framework

To establish asymptotic properties in sampling from a finite population, it is necessary to deal with sequences of different populations, vectors of design variables, design measures, samples and vectors of study variables.

Define $(N_\gamma)_{\gamma \in \mathbb{N}}$ a sequence of population sizes , such that $\lim_\gamma N_\gamma = +\infty$, and for $\gamma \in \mathbb{N}$, we define the γth population as the set $U_\gamma = \{1, \ldots, N_\gamma\}$. If not otherwise specified, all random variables are still defined on the probability space $(\Omega, \mathscr{A}, \mathrm{P})$. For all $\gamma \in \mathbb{N}$, we define two matrices of random variables: $\mathcal{Y}_\gamma = (Y_{\gamma k})_{k \in U_\gamma}$ and $\mathcal{Z}_\gamma = (Z_{\gamma k})_{k \in U_\gamma}$, where for $\gamma \in \mathbb{N}$, $k \in U_\gamma$, $Y_{\gamma k} : (\Omega, \mathscr{A}) \to (\mathscr{Y}, \mathscr{T}_Y)$ and $Z_{\gamma k} : (\Omega, \mathscr{A}) \to (\mathscr{Z}, \mathscr{T}_Z)$. The vector $Y_{\gamma k}$ corresponds to the study variables, and $Z_{\gamma k}$ to the design variables associated to the kth element of the population U_γ. Let Π_γ be a sequence of random design measures such that $\forall \gamma \in \mathbb{N}$:

$$\begin{cases} \Pi_\gamma : \Omega \to \mathbb{\Pi}_{N_\gamma}, \\ \exists D_\gamma \in \mathscr{F}\left(\mathscr{Z}^{N_\gamma}, \mathbb{\Pi}_\gamma\right) \text{ such that } \Pi_\gamma = D_\gamma(\mathcal{Z}_\gamma), \\ \forall z \in \mathscr{Z}^{N_\gamma},\ r \text{ a permutation of } U_\gamma,\ A \subset \mathbb{N}^{N_\gamma}\ \left(D_\gamma(z)\right)(A) = \left(D_\gamma(r.z)\right)(r.A). \end{cases}$$

For $\gamma \in \mathbb{N}$, define the random variable $\mathcal{I}_\gamma$ with value in $\mathbb{N}^{N_\gamma}$ such that

$$\begin{cases} \mathrm{P}^{\Pi_\gamma, \mathcal{Y}_\gamma, \mathcal{Z}_\gamma} - a.s.(\mathrm{p}, y, z), \mathrm{P}^{\mathcal{I}_\gamma | \Pi_\gamma = \mathrm{p}, \mathcal{Y}_\gamma = y, \mathcal{Z}_\gamma = z} = \mathrm{P}^{\mathcal{I} | \Pi = \mathrm{p}}, \\ \mathrm{P}^{\Pi_\gamma} - a.s.(\mathrm{p}),\ \mathrm{P}^{\mathcal{I}_\gamma | \Pi_\gamma = \mathrm{p}} = \mathrm{p}\,. \end{cases}$$

First and second order inclusion probabilities will be denoted

$$\pi_{\gamma,k} = \Pi_\gamma\left(\left\{ i \in \mathbb{N}^{N_\gamma} \middle| \, i_k \geqslant 1 \right\}\right),$$

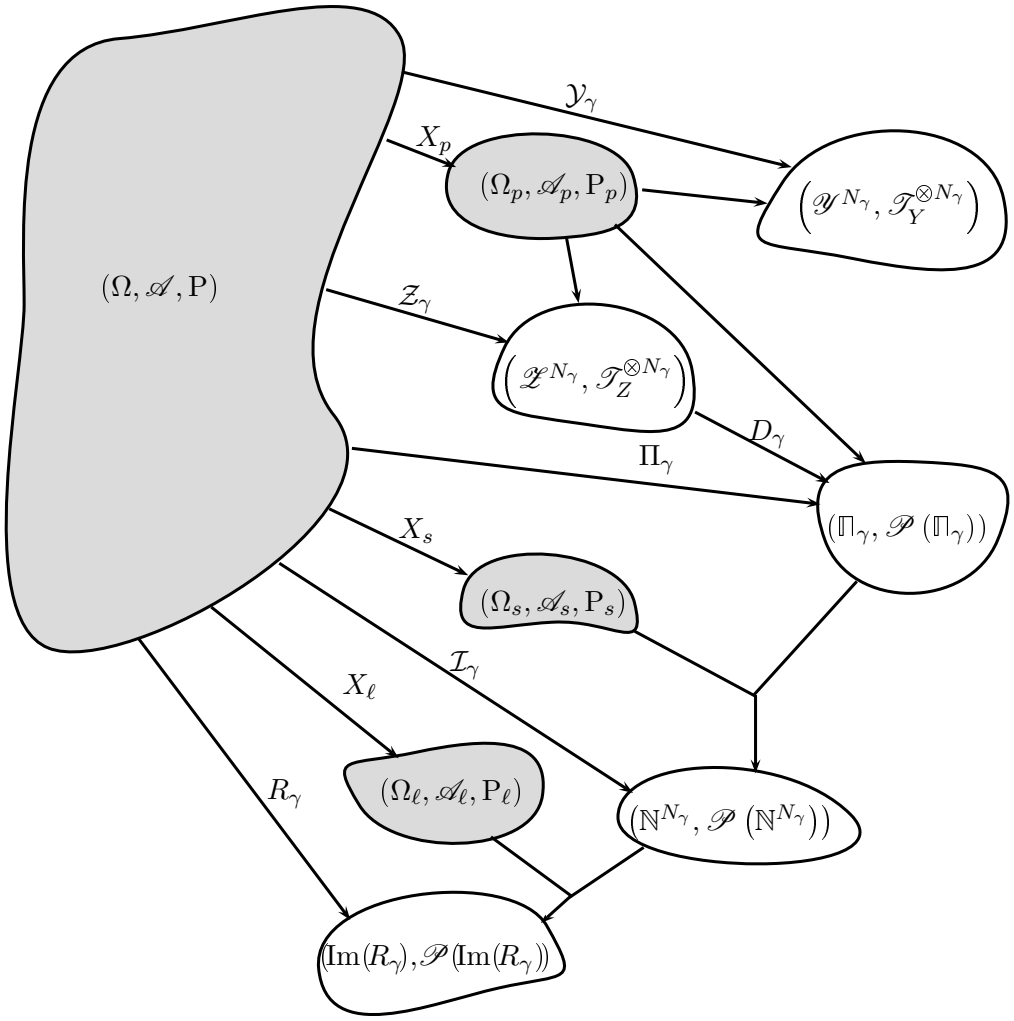
and

$$\pi_{\gamma,k,l} = \Pi_\gamma\left(\left\{ i \in \mathbb{N}^{N_\gamma} \middle| \, i_{\geqslant} 1, i_l \geqslant 1 \right\}\right)$$

For $\gamma \in \mathbb{N}$, define the sample size as the random variable $n_\gamma = \sum_{k=1}^{N_\gamma} I_{\gamma k}$. Define a random variable R_γ, that satisfies:

$$\begin{cases} R_\gamma(\omega) \in \left\{ r \in \mathscr{F}\left(\{1, \ldots, n_\gamma(\omega)\}, U_\gamma\right) | \forall k \in U_\gamma, \#\left(\{ l \in \{1, \ldots, n_\gamma(\omega)\} | r(l) = k \}\right) = I_{\gamma k}(\omega) \right\}, \\ \mathrm{P}^{R_\gamma | \mathcal{I}_\gamma = i, \mathcal{Y}_\gamma = y, \mathcal{Z}_\gamma = z} = \mathrm{P}^{R_\gamma | \mathcal{I}_\gamma = i}, \\ \mathrm{P}^{R_\gamma | \mathcal{I}_\gamma = i} \text{ is the uniform law on } \left\{ r \in U_\gamma^{\{1, \ldots, n_\gamma\}} \middle| \, \forall k \in U_\gamma, \#(\{ l \in \{1, \ldots, n_\gamma\} | \, r(l) = k \}) = i_k \right\}. \end{cases}$$

Figure 2.2: Commutative diagram for $\mathcal{Y}_\gamma, \mathcal{Z}_\gamma, \Pi_\gamma, \mathcal{I}_\gamma, R_\gamma$



The links between the random variables $\mathcal{Y}_\gamma, \mathcal{Z}_\gamma, \Pi_\gamma, \mathcal{I}_\gamma, R_\gamma$ are summarized by the commutative diagram of figure 2.2.

The observation is a random variable G_γ, that is a function of $\mathcal{I}_\gamma, \mathcal{Y}_\gamma, R_\gamma$:

$$G_\gamma = g_\gamma(\mathcal{I}_\gamma, \mathcal{Y}_\gamma, R_\gamma)$$

In all the following examples, for a complete description of the model, we will just need to specify the sequence of population models $\mathrm{P}^{\mathcal{Y}_\gamma, \mathcal{Z}_\gamma}$, the sequence of design measure functions D_γ, and the sequence of observation functions g_γ.

2.6 The sample pdf and the limit sample pdf

We consider the exchangeable population model. From now on, in order to consider weak convergence of the sample distribution, which is defined below, we restrict ourselves to the case where $\mathscr{Y} = \mathbb{R}^p$, and $\mathscr{T}_Y = \mathscr{B}_{\mathbb{R}^p}$, with $p \in \mathbb{N}\backslash\{0\}$,$\mathscr{B}_{\mathbb{R}^p}$ denoting the σ-algebra of Borel sets. We assume that $\forall \gamma \in \mathbb{N}$, $Y_{\gamma 1}$ has

a density f_γ with respect to a Radon positive measure μ_Y on $(\mathscr{Y}, \mathscr{T}_Y)$. We define the sample probability density function (pdf) as a weighted version of f_γ:

Definition 2.7. *The sample pdf*
Let $\gamma \in \mathbb{N}$. Assume $0 < \mathrm{E}[I_{\gamma 1}] < +\infty$. As $I_{\gamma k}$ is a positive random variable, we can define $\mathrm{E}[I_{\gamma k}|Y_{\gamma k} = y]$. The sample pdf is the function $\rho_\gamma f_\gamma$, where

$$\begin{aligned} \rho_\gamma : \quad \mathscr{Y} &\to \mathbb{R} \\ y &\mapsto \rho_\gamma(y) = \frac{\mathrm{E}[I_{\gamma k}|Y_{\gamma k} = y]}{\mathrm{E}[I_{\gamma k}]} \end{aligned}$$

and the sample probability measure is the measure $(\rho_\gamma f_\gamma).\mu_Y$.

Remark In Pfeffermann and Sverchkov (2009), the sample pdf is defined in the case of sampling without replacement, where $\mathscr{Y} = \mathbb{R}^2$, and is the function:

$$x, y \mapsto \left(\mathrm{P}(I_{\gamma k} = 1 \mid Y_{\gamma k} = y, X_{\gamma k} = x) / \mathrm{P}(I_k = 1 | X_{\gamma k} = x)\right) \left(d\,\mathrm{P}^{Y_{\gamma k}|X_{\gamma k}=x} / \mathrm{d}\lambda_p\right)(y),$$

where $\mathcal{X}_\gamma = (X_{\gamma k})_{k \in U_\gamma}$ is a vector of covariate variables, and the analyst knows at least $(X_{\gamma k})_{k \in U_\gamma, I_{\gamma k} \geqslant 1}$. In this dissertation, for simplicity, we do not consider explicitly the covariates in the first sections, but in our case, we can choose a dimension of $\mathscr{Y}$ that allows consideration of models with covariates. We extend the definition of the sample pdf to take into account the sampling with replacement case.

Property 2.6. *The weighted pdf $\rho_\gamma f$ is a probability density function.*

The following two properties deal with the distribution of the responses in the sample. In the case of sampling without replacement, the distribution of an observation can be characterized by its sample pdf, defined in Pfeffermann and Krieger (1992) as the conditional density of Y_k given $I_{\gamma k} = 1$. In the case of with replacement sampling, the distribution of any coordinate of a vector whose size is random (and can even be 0), cannot be defined without conditioning on the size of the sample. In both cases, the definition 2.7 of ρ_γ applies.

Property 2.7. *Relation to sample pdf in the case of fixed sized sampling. If we suppose that n_γ is strictly positive and not random then the sample pdf is the density of $Y_{\gamma R_\gamma(l)}$ with respect to μ_Y:*

$$\mathrm{P}^{Y_{R_\gamma(1)}} = (\rho_\gamma f_\gamma).\mu_Y,$$

and $\forall A \in \mathscr{T}_Y$, $\mathrm{E}\left[n_\gamma^{-1} \sum_{l=1}^{n_\gamma} \mathbb{1}_A\left(Y_{\gamma R_\gamma(l)}\right)\right] = \mathrm{P}^{Y_{\gamma R_\gamma(l)}}(A) = \mathrm{E}\left[\mathbb{1}_A(Y_{\gamma R_\gamma(1)})\right] = \left((\rho_\gamma f_\gamma).\mu_Y\right)(A)$, where $A \mapsto n^{-1} \sum_{l=1}^{n_\gamma} \mathbb{1}_A\left(Y_{\gamma R_\gamma(l)}\right)$ is the sample empirical measure of A.

Proof. See appendix B.1. □

Property 2.8. *Relation to sample pdf in the case of sampling without replacement.*

In the case of sampling without replacement, the sample pdf is the conditional pdf of $Y_{\gamma k}$ given that the kth element is selected:

$$\mathrm{P}^{Y_{\gamma k}|I_{\gamma k}=1} = (\rho_\gamma f_\gamma).\mu_Y.$$

Proof. By Bayes' rule,

$$\frac{\mathrm{dP}^{Y_{\gamma k}|I_{\gamma k}=1}}{\mathrm{d}\mu_Y}(y) = \frac{\mathrm{P}(I_{\gamma k} = 1 \mid Y_{\gamma k} = y) f_\gamma(y))}{\int \mathrm{P}(I_{\gamma k} = 1 \mid Y_{\gamma k} = y_1)\, f_\gamma \,\mathrm{d}\mu_Y} = \rho_\gamma(y) f_\gamma(y).$$

(See also (Pfeffermann and Sverchkov, 2009, eq.(1) p.457)) □

The following conditions will allow us to define the limit sample pdf.

•**A2.0.** *$\exists \mu_Y^*$ a Radon measure on $(\mathscr{Y}, \mathscr{T}_Y)$, and two measurable functions f_∞ and ρ_∞ from $(\mathscr{Y}, \mathscr{T}_Y)$ to $(\mathbb{R}^+, \mathscr{B}_{R^+})$, such that:*

$$\begin{cases} f_\gamma.\mu_Y \xrightarrow[\gamma\to\infty]{weakly} = f_\infty.\mu_Y^*, & \text{(2.0a)} \\ (\rho_\gamma f_\gamma).\mu_Y \xrightarrow[\gamma\to\infty]{weakly} (\rho_\infty f_\infty).\mu_Y^*. & \text{(2.0b)} \end{cases}$$

Remark The distribution $f_\infty.\mu_Y^*$ is the limit population distribution. In the case of the fixed population model, μ_Y is the counting measure and μ_Y^* can be different from μ_Y. In the case of the iid superpopulation model, $\exists f$ such that $\forall \gamma \in \mathbb{N}$, $f_\gamma = f$, and then $f_\infty.\mu_Y^* = f.\mu_Y$.

Definition 2.8. *Under A2.0, we define:*

$$\rho_\infty f_\infty$$

as the limit sample pdf and

$$(\rho_\infty.f_\infty).\mu_Y^*$$

as the limit sample distribution.

Conclusion

The specification of a model for responses that are the outcome of a survey requires the description of the population model, the design measure function, and the observation function. Under a very general population model, we propose a definition of non-informative selection, and give conditions where a limit sample pdf can be defined. Under some conditions, the weighted distribution $\rho_\infty.f_\infty\mu_Y^*$ may be used to approximate the law of the observation in a context of inference on the distribution of the responses in the population. We study some asymptotic properties of the probability laws of the sequence of observations $\mathrm{P}^{R_\gamma.\mathcal{Y}_\gamma|n=n^*}$ and compare them to those of the sequence of approximated probability laws $(\rho_\infty.f_\infty\mu_Y^*)^{\otimes n^*}$.

References

Cassel, C.-M., Särndal, C.-E., and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley-Interscience [John Wiley & Sons], New York. Wiley Series in Probability and Mathematical Statistics.

Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons Inc.

Gourieroux, C. (1981). *Théorie des Sondages*. Economica.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Pfeffermann, D. and Krieger, A. M. (1992). Maximum likelihood estimation for complex sample surveys. *Survey Methodology*, 18(2):225–239.

Pfeffermann, D. and Sverchkov, M. Y. (2009). Inference under Informative Sampling. In Pfefferman, D. and Rao, C., editors, *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of Statistics*, pages 455–487. Elsevier.

Skinner, C. J. (1994). Sample models and weights. In statistical association, A., editor, *Proceedings of the Section on Survey Research Methods*, pages 133–142, Washington, DC.

Chapter 3

Uniform convergence of the sample cdf

In this chapter we consider the case where Y is a random variable with a density with respect to the Lebesgue measure on $\mathbb{R}$, and establish conditions on the sequence of sampling schemes for the Glivenko-Cantelli theorem to hold in the case of informative selection.

3.1 Results

3.1.1 Asymptotic framework and assumptions

In this chapter the population model considered is a particular case of the exchangeable population model of section 2.3.3. Specifically, we assume that $(\mathscr{Y}, \mathscr{T}_Y) = (\mathbb{R}, \mathscr{B}_\mathbb{R})$ and that $\forall \gamma \in \mathbb{N}$, $k \in \{1, \dots, N_\gamma\}$, $Y_{\gamma k} = Y_k$ where $(Y_k)_{k\in\mathbb{N}}$ is a sequence of iid random variables, $Y_k : (\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}_\mathbb{R})$. We assume that Y_k admits a density f with respect to λ, the Lebesgue measure on $\mathbb{R}$.

Definition 3.1. *For $\gamma \in \mathbb{N}$, the empirical sample cdf is the random process $F_\gamma : \mathbb{R} \to [0,1]$ via*

$$F_\gamma(\alpha) = \frac{\sum_{k\in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_k) I_{\gamma k}}{\mathbb{1}_{\mathcal{I}_\gamma = 0} + n_\gamma}.$$

Definition 3.2. *Given γ, let $k, \ell \in U_\gamma$ with $k \neq \ell$. Let*

$$\begin{aligned}
m_\gamma(y) &= \mathrm{E}\left[I_{\gamma k} \mid Y_k = y\right] \\
v_\gamma(y) &= \mathrm{Var}\left[I_{\gamma k} \mid Y_k = y\right] \\
m'_\gamma(y_1, y_2) &= \mathrm{E}\left[I_{\gamma k} \mid Y_k = y_1, Y_\ell = y_2\right] \\
c_\gamma(y_1, y_2) &= \mathrm{Cov}\left[I_{\gamma k}, I_{\gamma \ell} \mid Y_k = y_1, Y_\ell = y_2\right] \\
d_\gamma(y_1, y_2) &= m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) m_\gamma(y_1) m_\gamma(y_2).
\end{aligned}$$

(These definitions do not depend on the choice of k, ℓ under the exchangeability conditions (2.4) and (2.8)).

We give conditions which, like A2.0, ensure the existence of a limit sample pdf:

•A3.0.

$$\begin{cases}
\exists M : (\mathbb{R}, \mathscr{B}_\mathbb{R}) \to \left(\mathbb{R}^+, \mathscr{B}_{\mathbb{R}^+}\right) \text{ measurable, such that } \begin{cases} \forall \gamma \in \mathbb{N}, m_\gamma < M \\ \int M f \, \mathrm{d}\lambda < \infty \end{cases} & \text{(3.0a)} \\
\exists m_\infty : (\mathbb{R}, \mathscr{B}_\mathbb{R}) \to \left(\mathbb{R}^+, \mathscr{B}_{\mathbb{R}^+}\right) \text{ measurable, such that } \begin{cases} \forall y \in \mathbb{R}, \lim_\gamma m_\gamma(y) = m_\infty(y) \\ \int m_\infty f \, \mathrm{d}\lambda > 0. \end{cases} & \text{(3.0b)}
\end{cases}$$

Definition 3.3. *Under A3.0, ρ_∞ is defined (see definition 2.8) and equals $\left(\int m_\infty f \, \mathrm{d}\lambda\right)^{-1} m_\infty$. Define the limit sample cdf*

$$F_\infty : \mathbb{R} \to [0,1],\ \alpha \mapsto \int \mathbb{1}_{]-\infty,\alpha]} \rho_\infty f \, \mathrm{d}\lambda.$$

The limit sample cdf may differ from the population cdf, defined as

$$F : \mathbb{R} \to [0,1], F : \alpha \mapsto F(\alpha) = \int \mathbb{1}_{]-\infty,\alpha]} f \, \mathrm{d}\lambda.$$

For a sequence of real positive numbers $(b_\gamma)_{\gamma\in\mathbb{N}}$, and a sequence of real numbers $(a_\gamma)_{\gamma\in\mathbb{N}}$ let $a = o_\gamma(b)$ denote $\forall \varepsilon \in]0,+\infty]\,,\ \exists \Gamma \in \mathbb{N}$ such that $\forall \gamma \in \mathbb{N}, [\gamma \geqslant \Gamma \Rightarrow |a_\gamma| < \varepsilon b_\gamma]$. In the next two assumptions, we define sufficient conditions for uniform L_2 convergence and uniform a.s. convergence of the empirical sample cdf.

•A3.1. Uniform L_2 convergence conditions:

$$\begin{cases} \int c_\gamma(y_1,y_2) f(y_1) f(y_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 = o_\gamma(1) & \text{(3.1a)} \\ \int d_\gamma(y_1,y_2) f(y_1) f(y_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 = o_\gamma(1) & \text{(3.1b)} \\ \int (v_\gamma + m_\gamma^2) f \, \mathrm{d}\lambda = o_\gamma(N_\gamma) & \text{(3.1c)} \\ \mathrm{P}\left(\{\mathcal{I}_\gamma = 0\}\right) = o_\gamma(1). & \text{(3.1d)} \end{cases}$$

•A3.2. Uniform almost sure convergence conditions:

$$\forall y \in \mathbb{R}^{\mathbb{N}} \text{ such that } \sup_{\alpha' \in \mathbb{R}} \left| \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha']}(y_k)}{N_\gamma} - \int \mathbb{1}_{(-\infty,\alpha']} f \, \mathrm{d}\lambda \right| = o_\gamma(1),$$

$$\begin{cases} \forall \alpha \in \mathbb{R},\ \mathrm{Var}\left[\sum_{k\in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) I_{\gamma k} \,\middle|\, \mathcal{Y}_\gamma = (y_1,\ldots,y_{N_\gamma}) \right] = o_\gamma(N_\gamma^2) & \text{(3.2a)} \\ \forall \alpha \in \mathbb{R},\ \sum_{k\in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) \left(\mathrm{E}\left[I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1 \ldots y_{N_\gamma}) \right] - m_\gamma(y_k) \right) = o_\gamma(N_\gamma) & \text{(3.2b)} \\ \mathrm{P}^{\mathcal{I}_\gamma = 0 | \mathcal{Y}_\gamma = (y_1 \ldots y_{N_\gamma})} = o_\gamma(1). & \text{(3.2c)} \end{cases}$$

Properties of sampling without replacement In the case of sampling without replacement, A3.0 and A3.1 can be replaced by a simpler set of sufficient conditions for uniform L_2 convergence.

•A3.3. Uniform L_2 convergence conditions under sampling without replacement:

$$\begin{cases} \exists\, m_\infty : \mathbb{R} \to \mathbb{R}^+ \ \lambda\text{-measurable s.t.} \begin{cases} m_\gamma \to m_\infty \text{ pointwise as } \gamma \to \infty \\ \int m f \, \mathrm{d}\lambda > 0 \end{cases} & \text{(3.3a)} \\ \forall y_1, y_2, c_\gamma(y_1,y_2) = o_\gamma(1) & \text{(3.3b)} \\ \forall y_1, y_2,\ m'_\gamma(y_1,y_2) - m_\gamma(y_2) = o_\gamma(1) & \text{(3.3c)} \\ \mathrm{P}(\mathcal{I}_\gamma = 0) = o_\gamma(1) & \text{(3.3d)} \\ \text{the selection is without replacement.} & \text{(3.3e)} \end{cases}$$

These conditions imply A3.0 and A3.1.

Proof. Since $I_{\gamma k} \in \{0,1\}$, A3.0a and A3.1c always hold. By applying the Lebesgue dominated convergence theorem, we obtain that A3.1a is verified when $\forall y_1, y_2, c_\gamma(y_1, y_2) = o_\gamma(1)$ and A3.1b is verified when $\forall y_1, y_2,\ m'_\gamma(y_1, y_2) - m_\gamma(y_2) = o_\gamma(1)$.

□

An important special case of sampling without replacement is non-informative selection, with $\mathcal{I}_\gamma$ independent of $\mathcal{Y}_\gamma$ for all $\gamma \in \mathbb{N}$. In this case, the sample obtained is an iid sample of size $n_\gamma = \sum_{k \in U_\gamma} I_{\gamma k}$ (Fuller, 2009, Thm. 1.3.1), and the classic Glivenko-Cantelli theorem can be applied as soon as $n_\gamma \xrightarrow{\text{a.s.}} \infty$ as $\gamma \to \infty$. The assumptions of Theorem 3.1 and Theorem 3.2 will then just ensure that the expectation of the sample size will grow to infinity, and that its variations are small enough to avoid very small samples. We can thus replace A3.0–A3.2 by a simpler set of sufficient conditions.

•A3.4. Uniform L_2 and a.s. convergence conditions under independent sampling without replacement:

$$\begin{cases} \mathcal{Y}_\gamma \textit{ and } \Pi_\gamma \textit{ are independent} \\ \textit{The selection is without replacement} \\ \lim_\gamma {N_\gamma}^{-1} \mathrm{E}\left[n_\gamma\right] = m \neq 0 \\ \mathrm{Var}\left[n_\gamma\right] = o_\gamma(N_\gamma^2). \end{cases} \tag{3.4}$$

A3.4 implies A3.0–A3.2.

Proof. We first show that A3.4 implies A3.3. Because $\mathcal{I}_\gamma$ and $\mathcal{Y}_\gamma$ are independent, the exchangeability conditions (2.4) and (2.8) imply $m_\gamma(y) = \mathrm{E}\left[I_{\gamma 1}\right] = {N_\gamma}^{-1} \mathrm{E}\left[n_\gamma\right]$ and ${N_\gamma}^{-1} \mathrm{E}\left[n_\gamma\right] \to m$ by A3.4, so A3.3a holds. Exchangeability also implies

$$\mathrm{E}\left[I_{\gamma 1} I_{\gamma 2}\right] = \frac{\sum_{k,\ell \in U_\gamma : k \neq \ell} \mathrm{E}\left[I_{\gamma k} I_{\gamma \ell}\right]}{N_\gamma(N_\gamma - 1)} = \mathrm{E}\left[\frac{\sum_{k,\ell \in U_\gamma : k \neq \ell} I_{\gamma k} I_{\gamma \ell}}{N_\gamma(N_\gamma - 1)}\right] = \mathrm{E}\left[\frac{n_\gamma(n_\gamma - 1)}{N_\gamma(N_\gamma - 1)}\right]$$

so

$$c_\gamma(y_1, y_2) = \mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2}\right] = \mathrm{E}\left[\frac{n_\gamma(n_\gamma - N_\gamma)}{N_\gamma^2(N_\gamma - 1)}\right] + \mathrm{Var}\left[\frac{n_\gamma}{N_\gamma}\right] = o_\gamma(1) \tag{3.5}$$

by A3.4, so A3.3 is obtained, and A3.4 holds by independence. Finally,

$$\begin{aligned} \mathrm{P}\left(n_\gamma = 0\right) = \mathrm{P}\left(n_\gamma < 1\right) &= \mathrm{P}\left(n_\gamma - \mathrm{E}\left[n_\gamma\right] < 1 - \mathrm{E}\left[n_\gamma\right]\right) \\ &\leqslant \mathrm{P}\left(\left|n_\gamma - \mathrm{E}\left[n_\gamma\right]\right| > \mathrm{E}\left[n_\gamma\right] - 1\right) \\ &\leqslant \frac{\mathrm{Var}\left[n_\gamma\right]}{\left(\mathrm{E}\left[n_\gamma\right] - 1\right)^2} = o_\gamma(1), \end{aligned} \tag{3.6}$$

establishing A3.3e.

We next show that A3.4 implies A3.2. For all $\alpha \in \mathbb{R}$,

$$\begin{aligned} &\mathrm{Var}\left[\sum_{k \in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_k) I_{\gamma k} \,\middle|\, \mathcal{Y}_\gamma = (y_1 \dots y_{N_\gamma})\right] \\ &= \sum_{k \in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k)\, \mathrm{Var}\left[I_{\gamma k}\right] \\ &\quad + \sum_{k,\ell \in U_\gamma : k \neq \ell} \mathbb{1}_{]-\infty,\alpha]}(y_k) \mathbb{1}_{]-\infty,\alpha]}(y_\ell)\, \mathrm{Cov}\left[I_{\gamma k}, I_{\gamma \ell}\right] \\ &\leqslant N_\gamma + N_\gamma(N_\gamma - 1) o_\gamma(1) = o_\gamma\left(N_\gamma^2\right) \end{aligned}$$

by equation (3.5), so A3.2a holds. By independence,

$$\mathrm{E}\left[I_{\gamma k} \mid \mathcal{Y}_\gamma = (y_1, \ldots, y_{N_\gamma})\right] = \mathrm{E}\left[I_{\gamma k} \mid Y_k = y_k\right] = m_\gamma(y_k),$$

so A3.2b holds. Finally,

$$\mathrm{P}\left(\mathcal{I}_\gamma = (0, \ldots, 0) | \mathcal{Y}_\gamma = (y_1, \ldots, y_{N_\gamma})\right) = \mathrm{P}\left(n_\gamma = 0\right) = o_\gamma(1)$$

by independence and (3.6), so A3.2c holds.

□

We have seen (see Example 2.6) that if $\mathcal{Y}_\gamma$ and $\mathcal{Z}_\gamma$ are independent, then $\mathcal{Y}_\gamma$ and Π_γ are independent, and $\mathcal{Y}_\gamma$ and $\mathcal{I}_\gamma$ are also independent. If in addition the sample is without replacement, then $\mathrm{P}^{R.\mathcal{Y}_\gamma | n_\gamma = n_0} = (f.\lambda)^{\otimes n_0}$. In that case, we also notice that $F = F_\infty$ and $\forall n^* \in \{1, \ldots, N_\gamma\}, \mathrm{E}\left[F_\gamma | n_\gamma = n^*\right] = F$.

Remark

In conventional finite population asymptotics (Breidt and Opsomer, 2000, 2008; Isaki and Fuller, 1982; Robinson and Särndal, 1983), conditions on design covariances $\mathrm{Cov}\left[I_{\gamma k}, I_{\gamma \ell}\right]$ are imposed to guarantee that the Horvitz-Thompson estimator $\sum_{k \in U_\gamma} y_k I_{\gamma k} (\mathrm{E}\left[I_{\gamma k}\right])^{-1}$ is consistent. Typically, these conditions imply that the variance of the Horvitz-Thompson estimator is $O_\gamma(N_\gamma^2/(N_\gamma \pi_{*\gamma}))$, where $N_\gamma \pi_{*\gamma} \to \infty$ is a sequence of lower bounds on the expected sample size, $\mathrm{E}\left[n_\gamma\right]$. These same conditions can be used to show that $\mathrm{Var}\left[n_\gamma\right] = O_\gamma(N_\gamma^2/(N_\gamma \pi_{*\gamma})) = o_\gamma(N_\gamma^2)$, agreeing with A3.4.

3.1.2 Uniform convergence of the empirical sample cdf

In this section, we state the main results of the chapter: uniform L_2 convergence of the empirical sample cdf and uniform almost sure convergence of the empirical sample cdf. Important corollaries yield uniform convergence of sample quantiles on compact sets. Proofs are given in Appendix C.

3.1.2.1 Uniform L_2 convergence of the empirical sample cdf

Theorem 3.1. *Under A3.0 and A3.1, the empirical sample cdf converges uniformly in L_2 in the sense that*

$$\sup_{\alpha \in \mathbb{R}} |F_\gamma(\alpha) - F_\infty(\alpha)| = \|F_\gamma - F_\infty\|_\infty \xrightarrow[\gamma \to \infty]{L_2} 0.$$

Proof. See Appendix C.1.1. □

Definition 3.4. *The limit quantiles $\zeta_\infty :]0,1[\to \mathbb{R}$ are given by*

$$\zeta_\infty(x) = \inf\{y \in \mathbb{R} : F_\infty(y) \geqslant x\}$$

and the empirical quantiles $\zeta_\gamma :]0,1[\to \mathbb{R}$ are given by

$$\zeta_\gamma(x) = \inf\{y \in \mathbb{R} : F_\gamma(y) \geqslant x\}.$$

With this definition, we have the following corollary:

Corollary 3.1. *Suppose that F_∞ is continuous on $\mathbb{R}$ and $0 < F_\infty(y_1) = F_\infty(y_2) < 1 \Rightarrow y_1 = y_2$. Then, under A3.0 and A3.1, the empirical quantiles converge uniformly in probability to the limit quantiles,*

$$\sup_{x\in K} |\zeta_\gamma(x) - \zeta_\infty(x)| \xrightarrow[\gamma\to\infty]{P} 0$$

for all K a compact subset of $(0,1)$. Under the further hypothesis that f has compact support, the convergence is uniform in L_2:

$$\sup_{x\in K} |\zeta_\gamma(x) - \zeta_\infty(x)| \xrightarrow[\gamma\to\infty]{L_2} 0.$$

Proof. See Appendix C.2.1. □

3.1.2.2 Uniform almost sure convergence of the empirical cdf

The Glivenko-Cantelli theorem gives uniform almost sure convergence of the empirical sample cdf under iid sampling. We now consider uniform almost sure convergence under dependent sampling satisfying the second-order conditions of A3.2.

Asymptotic arguments in survey sampling consist first in embedding a specific sample scheme in a sequence of sample schemes. In the proof of the following representation theorem, we link the elements of the sequence of sample schemes in a way that ensures uniform almost sure convergence of the empirical cdf. We stress that in our result the vector of responses for the population remains the original $\mathcal{Y}_\gamma = (Y_k)_{k\in U_\gamma}$, and not another set of identically distributed random variables.

Theorem 3.2. *Under A3.0 and A3.2, there exist sequences of random variables $(I'_{\gamma k})_{\gamma\in\mathbb{N},k\in U_\gamma}$, $(Y'_k)_{k\in\mathbb{N}}$ defined on the probability space $\left(\Omega\times[0,1], \mathscr{A}\otimes\mathscr{B}_{[0,1]}, \mathrm{P}' = \mathrm{P}\otimes\lambda_{[0,1]}\right)$ such that*

- *$\|F'_\gamma - F_\infty\|_\infty$ converges P'-a.s. to 0*
- *$\forall\gamma\in\mathbb{N}$, $(\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma)$ and $(\mathcal{I}_\gamma, \mathcal{Y}_\gamma)$ have the same law*
- *$\forall\gamma\in\mathbb{N}$, $\omega\in\Omega$, $x\in[0,1]$, $\mathcal{Y}'_\gamma(\omega, x) = \mathcal{Y}_\gamma(\omega)$*

where $\mathscr{B}_{[0,1]}$ is the σ-field of Borel sets, $\lambda_{[0,1]}$ is the Lebesgue measure on $[0,1]$, $\mathcal{I}'_\gamma = \left(I'_{\gamma 1}, \ldots, I'_{\gamma N_\gamma}\right)$, $\mathcal{Y}'_\gamma = \left(Y'_1, \ldots, Y'_{N_\gamma}\right)$ and $F'_\gamma : \mathbb{R}\to[0,1]$ via

$$F'_\gamma(\alpha) = \frac{\sum_{k\in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y'_{\gamma k}) I'_{\gamma k}}{\sum_{k\in U_\gamma} I'_{\gamma k} + \mathbb{1}_{I'_\gamma = 0}}. \tag{3.7}$$

Proof. See appendix C.1.2. □

Corollary 3.2. *Suppose that F_∞ is continuous and $0 < F_\infty(y_1) = F_\infty(y_2) < 1 \Rightarrow y_1 = y_2$. If A3.0 and A3.2 hold, then for $(I'_{\gamma k})_{\gamma\in\mathbb{N},k\in U_\gamma}$ and $(Y'_k)_{k\in\mathbb{N}}$ that satisfy the conditions of Theorem 3.2, the empirical quantiles*

$$\zeta'_\gamma(x) = \inf\{y\in\mathbb{R} : F'_\gamma(y) \geqslant x\}$$

converge uniformly almost surely,

$$\sup_{x\in K} |\zeta_\gamma(x) - \zeta_\infty(x)| \xrightarrow[\gamma\to\infty]{a.s.} 0$$

for all K a compact subset of $(0,1)$.

Proof. See appendix C.2.2. □

3.2 Examples

We now consider a series of examples of selection mechanisms, motivated by real problems in surveys and other observational studies. We give examples where conditions A3.0, A3.1, A3.2 hold and where they fail.

3.2.1 Non-informative selection without replacement

- For any sequence of **fixed-size without-replacement designs** with $\mathcal{I}_\gamma$ independent of $\mathcal{Y}_\gamma$ (e.g., simple random sampling, stratified sampling with stratification variables independent of $\mathcal{Y}_\gamma$, rejective sampling (Hájek, 1981) with inclusion probabilities independent of $\mathcal{Y}_\gamma$, etc.), the condition A3.4 holds provided that $n_\gamma N_\gamma^{-1}$ converges to a strictly positive sampling rate.

  Proof. When $\mathcal{I}_\gamma$ and $\mathcal{Y}_\gamma$ are independent, then with the exchangeability conditions (2.4) and (2.8), $m_\gamma(y) = E[I_{\gamma 1}] = N_\gamma^{-1} E[n_\gamma]$, and assumption (3.4) becomes :

$$\begin{cases} \lim_{\gamma\to\infty} \mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2}\right] = o_\gamma(1), & (3.8) \\ \exists m \in \mathbb{R}_+^* \text{ s.t. } \lim_{\gamma\to\infty} \dfrac{E[n_\gamma]}{N_\gamma} = m. & (3.9) \end{cases}$$

  In addition,

$$\begin{aligned} \mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2}\right] &= \mathrm{E}\left[\frac{n_\gamma}{N_\gamma}\left(\frac{n_\gamma - 1}{N_\gamma - 1} - \frac{n_\gamma}{N_\gamma}\right)\right] + \mathrm{Var}\left[\frac{n_\gamma}{N_\gamma}\right] \\ &= \mathrm{E}\left[\frac{n_\gamma(n_\gamma - N_\gamma)}{N_\gamma^2(N_\gamma - 1)}\right] + V\left[\frac{n_\gamma}{N_\gamma}\right] \\ &= O_\gamma\left(\frac{1}{N_\gamma}\right) + \mathrm{Var}\left[\frac{n_\gamma}{N_\gamma}\right]. \end{aligned}$$

  So assumption (3.4) is verified when

$$\begin{cases} \exists m \in \mathbb{R}_+^* \text{ s.t. } \lim_{\gamma\to\infty} \dfrac{E[n_\gamma]}{N_\gamma} = m, & (3.10) \\ \lim \dfrac{\mathrm{Var}\left[n_\gamma\right]}{N_\gamma^2} = 0. & (3.11) \end{cases}$$

  The condition (3.10) ensures that the expectation of the sample size will grow to infinity, and the condition (3.11), that its variations are small enough to avoid very small samples. □

- Consider a sequence of **Bernoulli samples** with parameter $\alpha \in]0, 1]$. For $N \in \mathbb{N}\backslash\{0\}$, the Bernoulli design measure on $\mathbb{N}^N$ with parameter α is denoted $\mathrm{Bern}_{N,\alpha}$ and is characterized by:

$$\forall i \in \mathbb{N}^N,\ \mathrm{Bern}_{N,\alpha}(\{i\}) = \prod_{k=1}^{N} \mathbb{1}_{\{0,1\}}(i_k)\alpha^{i_k}(1-\alpha)^{1-i_k}.$$

  Assume $\forall z \in \mathscr{Z}^{N_\gamma}$, $D_\gamma(z) = \mathrm{Bern}_{N_\gamma,\alpha}$. Then $I_{\gamma 1}, \ldots, I_{\gamma N_\gamma}$ are iid Bernoulli(α) random variables, independent from $\mathcal{Y}$. Furthermore, $\mathrm{Var}\left[n_\gamma\right] = N_\gamma \alpha(1-\alpha)$ and condition A3.4 holds.

- **Poisson sampling** corresponds to a design in which $\mathscr{Z} = [0,1]$, $\mathcal{Z}_\gamma$ and $\mathcal{Y}_\gamma$ are independent, and $\forall z \in \mathscr{Z}^{N_\gamma}$, $D(z) = \text{Pois}_z$. In this case, $\pi_{\gamma k} = Z_{\gamma k}$ and the variance of n_γ is given by

$$\text{Var}\left[n_\gamma\right] = \sum_{k \in U_\gamma} \text{E}\left[Z_{\gamma k}(1 - Z_{\gamma k})\right] + \text{Var}\left[\sum_{k \in U_\gamma} Z_{\gamma k}\right].$$

  Note that the first term in this expression is always $o_\gamma(N_\gamma^2)$, so it suffices to consider the second.

  - In the case where the vector $[Z_{\gamma k}]_{k \in U_\gamma}$ is just a random permutation of a non-random vector $[z_{\gamma k}]_{k \in U_\gamma}$ (e.g, $\text{P}_\gamma^{\mathcal{Z}} = (N_\gamma!)^{-1} \sum_{r \in \mathfrak{S}_{N_\gamma}} \delta_{r.z_\gamma}$) then $\text{Var}\left[\sum_{k \in U_\gamma} Z_{\gamma k}\right] = \text{Var}\left[\sum_{k \in U_\gamma} z_{\gamma k}\right] = 0$ and A3.4 is satisfied when ${N_\gamma}^{-1} \sum_{k \in U_\gamma} z_{\gamma k}$ converges to a non-zero constant.
  - Assume there exists a fixed sequence $(n_\gamma^*)_{\gamma \in \mathbb{N}} \in \mathbb{N}^N$ such that $\forall \gamma \in \mathbb{N}$, $\text{P}-(a.s.)$, $\sum_{k \in U_\gamma} Z_{\gamma k} = n_\gamma^*$. Then $\text{Var}\left[\sum_{k \in U_\gamma} Z_{\gamma k}\right] = 0$ and A3.4 is satisfied when ${N_\gamma}^{-1} n_\gamma^*$ converges to a non-zero constant.
  - Let $a_\gamma, b_\gamma \in (0,1]$ with $a_\gamma \neq b_\gamma$. If

$$\left(Z_{\gamma 1} \ldots Z_{\gamma N_\gamma}\right) = \begin{cases} (a_\gamma \ldots a_\gamma) & \text{with probability } 1/2, \\ (b_\gamma \ldots b_\gamma) & \text{with probability } 1/2, \end{cases}$$

    then

$$\text{Var}\left[\sum_{k \in U_\gamma} Z_{\gamma k}\right] = N_\gamma^2 \frac{(a_\gamma - b_\gamma)^2}{4} \neq o_\gamma(N_\gamma^2).$$

    Then A3.4 is not verified and in fact if $N_\gamma a_\gamma = O_\gamma(1)$ we do not have uniform convergence of the empirical sample cdf.

3.2.2 Length-biased sampling

Length-biased sampling, in which $\text{P}(I_{\gamma k} = 1 \mid Y_k = y_k) = m_\gamma(y_k) \propto y_k$, is pervasive in real surveys and observational studies. Cox (1969) gives a now-classic example of sampling fibers in textile manufacture, in which $m_\gamma(y_k) \propto y_k$ = fiber length. In surveys of wildlife abundance, "visibility bias" means that larger individuals or groups are more noticeable (e.g., Patil and Rao, 1978), so $m_\gamma(y_k) \propto y_k$ = size of individual or group. "On-site surveys" are sometimes used to study people engaged in some activity like shopping in a mall (Nowell and Stanley, 1991) or fishing at the seashore (Sullivan et al., 2006); the longer they spend doing the activity, the more likely the field staff are to intercept and interview them, so $m_\gamma(y_k) \propto y_k$ = activity time. In mark-recapture surveys of wildlife populations, individuals that live longer are more likely to be recaptured, so $m_\gamma(y_k) \propto y_k$ = lifetime (e.g., Leigh, 1988). Similarly, in epidemiological studies of latent diseases, individuals who become symptomatic seek treatment and drop out of eligibility for sampling, while those with long latency periods are more likely to be sampled: $m_\gamma(y_k) \propto y_k$ = latency period. Finally, propensity to respond to a survey is often related to a variable of interest; e.g., higher response rates from higher-income individuals.

Suppose that f has compact, positive support: $\int \mathbb{1}_{[\epsilon, M]} f \, \text{d}\lambda = 1$ for some $0 < \epsilon < M < \infty$. For the γth finite population, consider Poisson sampling with inclusion probability proportional to Y, in the sense that $\{I_{\gamma k}\}_{k \in U_\gamma}$ are independent binary random variables, with

$$\text{P}\left(I_{\gamma k} = 1 \mid Y_k = y_k\right) = 1 - \text{P}\left(I_{\gamma k} = 0 \mid Y_k = y_k\right) = m_\gamma(y_k) \propto y_k.$$

Let $\tau_\gamma = y_k^{-1} \mathrm{P}\left(I_{\gamma k} = 1 \mid Y_k = y_k\right)$ be the common proportionality constant (independent of k), and assume that $\tau_\gamma \to \tau \in (0, M^{-1}]$ as $\gamma \to \infty$. Then

$$\begin{aligned} m_\gamma(y) = \tau_\gamma y \quad &\to \quad \tau y = m(y) \\ c_\gamma(y_k, y_\ell) = 0, \quad &\quad m'_\gamma(y_k, y_\ell) - m_\gamma(y_k) = 0 \\ \mathrm{P}\left(\mathcal{I}_\gamma = (0, \ldots, 0)\right) \quad &= \quad \mathrm{E}\left[\prod_{k \in U_\gamma} (1 - \tau_\gamma y_k)\right] \\ &\leqslant \quad (1 - \tau_\gamma \epsilon)^{N_\gamma} = \exp(N_\gamma \ln(1 - \tau_\gamma \epsilon)) = o_\gamma(1), \end{aligned}$$

so that A3.3 is verified. It then follows that the limiting cdf is given by

$$F_\infty(\alpha) = \int \mathbb{1}_{]-\infty, \alpha]} \frac{y}{\mathrm{E}\left[Y_1\right]} f \, \mathrm{d}\lambda. \tag{3.12}$$

3.2.3 Cluster sampling

Here is presented an example of non convergence to the limit sample cdf: the limit sample cdf F_∞ may exist (A3.0 holds), but F_γ fails to converge to it (A3.1, A3.2 fail). Suppose Y is uniform on $[0, 1]$. Assume $\mathcal{Z} = \mathcal{Y}$, and for a vector $z \in \mathbb{R}^N$, define $\zeta_\alpha(z)$ as the quantile of order α of the values of the vector z: $\zeta_\alpha(z) = \inf\left\{z_k \,\middle|\, k \in \{1, \ldots, N\}, N^{-1} \# \left\{\ell \in \{1, \ldots, N\} \,|\, z_\ell \leqslant z_k\right\} \geqslant \alpha\right\}$. Assume the design measure function is the function characterized by:

$\forall z \in [0, 1]^{N_\gamma}, i \in \mathbb{N}^{N_\gamma}$,

$$\begin{aligned} \left(D_\gamma(z)\right) \quad : \quad i \mapsto \frac{1}{2} \Bigg(&\left(\prod_{k=1}^{N_\gamma} \left(\mathbb{1}_{[0, \zeta_{0.2}(z)]}\left(z_k\right)\right)^{i_k} \left(1 - \mathbb{1}_{[0, \zeta_{0.2}(z)]}\left(z_k\right)\right)^{1 - i_k} \right) \\ &+ \left(\prod_{k=1}^{N_\gamma} \left(\mathbb{1}_{[\zeta_{0.8}(z), 1]}\left(z_k\right)\right)^{i_k} \left(1 - \mathbb{1}_{[\zeta_{0.8}(z), 1]}\left(z_k\right)\right)^{1 - i_k} \right) \Bigg). \end{aligned} \tag{3.13}$$

Equation (3.13) means that with probability $1/2$ the sample consisting of the elements that correspond to the smallest 20% of the values of $\mathcal{Y}_\gamma$ are selected, and with probability $1/2$ the sample consisting of the elements that correspond to the largest 20% of the values of $\mathcal{Y}_\gamma$ are selected.

Note that

$$\begin{aligned} \mathrm{Cov}\left[I_{\gamma k}, I_{\gamma \ell} \mid Y_k = y_1, Y_\ell = y_2\right] \quad &= \quad \frac{1}{2} \mathbb{1}_{]-\infty, 0.2]}(y_1) \mathbb{1}_{]-\infty, 0.2]}(y_2) \\ &\quad + \frac{1}{2} \mathbb{1}_{]0.2, \infty[}(y_1) \mathbb{1}_{]0.2, \infty[}(y_2) - \frac{1}{4} \end{aligned}$$

so that

$$\begin{aligned} \int c_\gamma(y_1, y_2) f(y_1) f(y_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 \quad &= \quad \frac{1}{2} F^2(\tau) + \frac{1}{2}(1 - F(\tau))^2 - \frac{1}{4} \\ &\neq \quad o_\gamma(1), \end{aligned}$$

and A3.1 fails to hold.

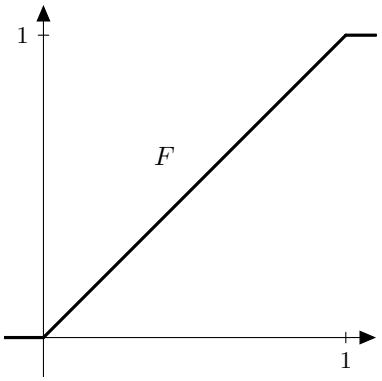
We plot the superpopulation cdf F in Figure 3.2(a) and the limit sample cdf F_∞ in Figure 3.2(b). By simulation, we generate $\mathcal{Y}_\gamma$ and $\mathcal{Z}_\gamma$ with $N_\gamma = 50$, according to the population model. We plot the population empirical cdf (the function: $\alpha \mapsto N_\gamma^{-1} \sum_{k \in U_\gamma} \mathbb{1}_{]-\infty, \alpha]}(Y_{\gamma k})$) in Figure 3.2(c). Then we draw two

independent samples, each according to the design measure $\Pi_\gamma(\mathcal{Z}_\gamma)$. It occurs that the first sample drawn contains the elements corresponding to the smallest 20% of the response values, and the second to the elements corresponding to the largest 20% of the response values. For both samples, we plot the sampled units (Figures 3.2(d), 3.2(f)) and the empirical sample cdf (Figures 3.2(e), 3.2(g)). This example can be regarded as a "worst-case" cluster sample: the sample consists of many elements but only one cluster, and the population is made up of a small number of large clusters, none of which is fully representative of the population.

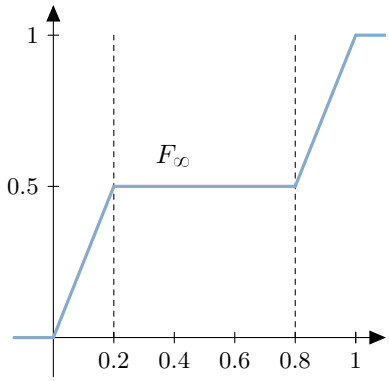
Remark If we had taken 50% instead of 20%, we would have had $F_\infty = F$, but even in that case, F_γ would have failed to converge to F_∞.

Figure 3.1: Cluster sampling example showing population cdf, limit sample cdf, empirical population cdf, and empirical sample cdf.

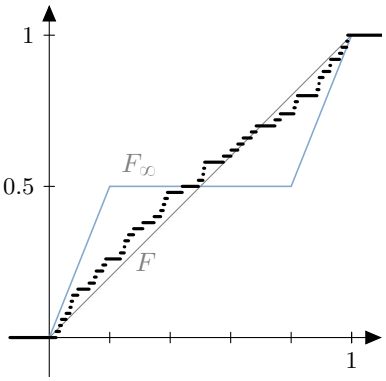
$N_\gamma = 50$, $n_\gamma = 10$, $\mathcal{Z}_\gamma = \mathcal{Y}_\gamma$



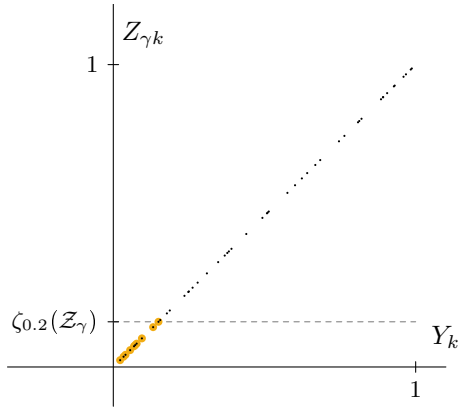
(a) Plot of the population cdf F



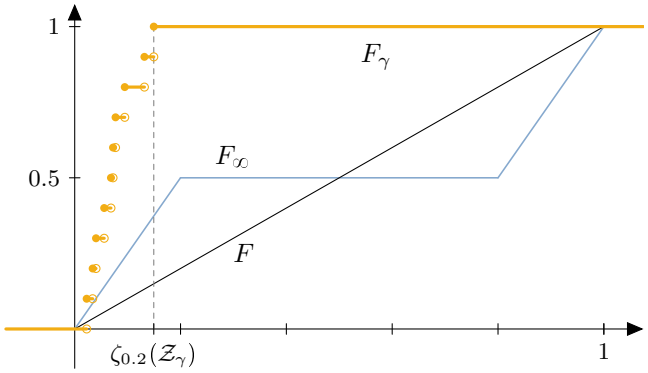
(b) Plot of the sample cdf F_∞



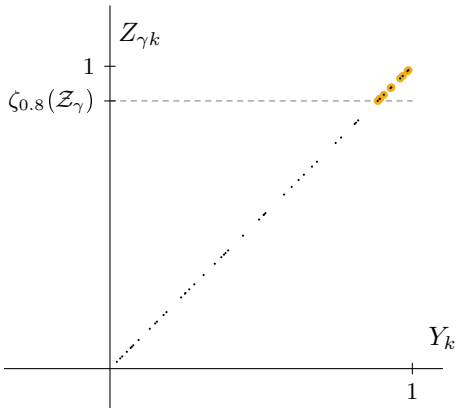
(c) Plot of the empirical population cdf.



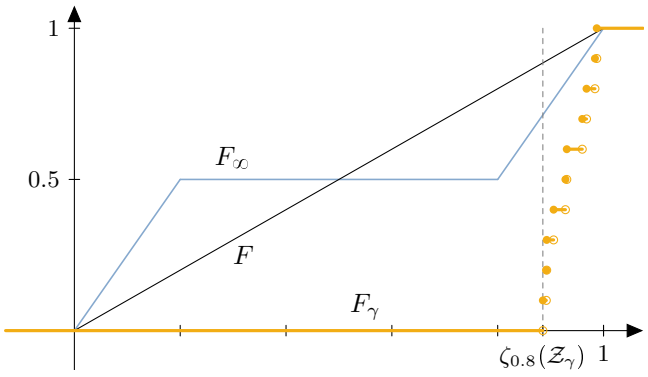
(d) First draw. Plot of $(Y_k, Z_{\gamma k})_{k \in U_\gamma}$. The large circles correspond to units selected, $N = 50$



(e) First draw. Plot of F_γ



(f) Second draw. Plot of $(Y_k, Z_{\gamma k})_{k \in U_\gamma}$. The large circles correspond to units selected, $N = 50$



(g) Second draw. Plot of F_γ

3.2.4 Cut-off sampling and take-all strata

In cut-off sampling a part of the population is excluded from sampling, so that $I_{\gamma k} = 0$ with probability one for some subset of U_γ. This may be due to physical limitations of the sampling apparatus, like a net that lets small animals escape, or may be due to a deliberate design decision. For example, a statistical agency may be willing to accept the bias inherent in cutting off small y-values if the y-distribution is highly skewed, as is often the case in establishment surveys (e.g., Särndal et al., 1992, §14.4).

Consider cut-off sampling with $I_{\gamma k} = 0$ for $\{k \in U_\gamma : y_k \leqslant \tau\}$, and simple random sampling without replacement of size $\min\{n_\gamma, N_\gamma - \sum_{j \in U_\gamma} \mathbb{1}_{(-\infty,\tau]}(y_j)\}$ from the remaining population, $\{j \in U_\gamma : y_j > \tau\}$.

Define $Z_k = \mathbb{1}_{(-\infty,\tau]}(Y_k)$ with corresponding realization $z_k = \mathbb{1}_{(-\infty,\tau]}(y_k)$. Let $\rho_\gamma = N_\gamma^{-1} n_\gamma$ and assume that $\lim_{\gamma\to\infty} \rho_\gamma = \rho$. We now verify A3.3.

Define $S_{\gamma A} = \sum_{j \in U_\gamma : j \notin A} Z_j$. By the weak law of large numbers, $N_\gamma^{-1} S_{\gamma A} \xrightarrow{P} F(\tau)$ as $\gamma \to \infty$ for $A = \{k\}$ or $A = \{k, \ell\}$, and so for those sets A we have

$$\begin{aligned}\lim_{\gamma\to\infty} \mathrm{E}\left[\frac{\rho_\gamma - {N_\gamma}^{-1} S_{\gamma A}}{1 - N_\gamma^{-1} S_{\gamma A}} \mathbb{1}_{\{\rho_\gamma > {N_\gamma}^{-1} S_{\gamma A}\}}\right] \\ = \frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}\end{aligned}$$

by the uniform integrability of the integrand. With the same argument,

$$\begin{aligned}\lim_{\gamma\to\infty} \mathrm{E}\left[\frac{\left(n_\gamma - S_{\gamma\{k,\ell\}}\right)\left(n_\gamma - 1 - S_{\gamma\{k,\ell\}}\right)}{\left(N_\gamma - S_{\gamma\{k,\ell\}}\right)\left(N_\gamma - 1 - S_{\gamma\{k,\ell\}}\right)} \mathbb{1}_{\{n_\gamma > S_{\gamma\{k,\ell\}}\}}\right] \\ = \left(\frac{\rho - F(\tau)}{1 - F(\tau)}\right)^2 \mathbb{1}_{\{\rho > F(\tau)\}}\end{aligned}$$

Using conditional first and second-order inclusion probabilities under simple random sampling, we have

$$\begin{aligned}m_\gamma(y_k) &= z_k + (1 - z_k)\,\mathrm{E}\left[\frac{n_\gamma - S_{\gamma\{k\}}}{N_\gamma - S_{\gamma\{k\}}} \mathbb{1}_{\{n_\gamma > S_{\gamma\{k\}}\}}\right] \\ &\to z_k + (1 - z_k)\frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}\end{aligned}$$

$$\begin{aligned}m'_\gamma(y_\ell, y_k) &= z_k + (1 - z_\ell)(1 - z_k)\,\mathrm{E}\left[\frac{n_\gamma - S_{\gamma\{k,\ell\}}}{N_\gamma - S_{\gamma\{k,\ell\}}} \mathbb{1}_{\{n_\gamma > S_{\gamma\{k,\ell\}}\}}\right] \\ &\quad + z_\ell(1 - z_k)\,\mathrm{E}\left[\frac{n_\gamma - 1 - S_{\gamma\{k,\ell\}}}{N_\gamma - 1 - S_{\gamma\{k,\ell\}}} \mathbb{1}_{\{n_\gamma - 1 > S_{\gamma\{k,\ell\}}\}} \mathbb{1}_{\{N_\gamma - 1 > S_{\gamma\{k,\ell\}}\}}\right] \\ &\to z_k + (1 - z_k)\frac{(\rho - F(\tau))\mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)}\end{aligned}$$

$$
\begin{aligned}
d_\gamma(y_k, y_\ell) &= \mathrm{E}\left[I_{\gamma k} I_{\gamma \ell} \mid Y_k = y_k, Y_\ell = y_\ell\right] \\
&= z_k z_\ell + \{z_k(1 - z_\ell) + (1 - z_k) z_\ell\} \, \mathrm{E}\left[\frac{n_\gamma - 1 - S_{\gamma\{k,\ell\}}}{N_\gamma - 1 - S_{\gamma\{k,\ell\}}} \mathbb{1}_{\{n_\gamma - 1 > S_{\gamma\{k,\ell\}}\}}\right] \\
&\quad + (1 - z_k)(1 - z_\ell) \, \mathrm{E}\left[\frac{\left(n_\gamma - S_{\gamma\{k,\ell\}}\right)\left(n_\gamma - 1 - S_{\gamma\{k,\ell\}}\right)}{\left(N_\gamma - S_{\gamma\{k,\ell\}}\right)\left(N_\gamma - 1 - S_{\gamma\{k,\ell\}}\right)} \mathbb{1}_{\{n_\gamma > S_{\gamma\{k,\ell\}}\}}\right] \\
&\to z_k z_\ell + (1 - z_k)(1 - z_\ell) \left(\frac{\rho - F(\tau)}{1 - F(\tau)}\right)^2 \mathbb{1}_{\{\rho > F(\tau)\}} \\
&\quad + \{z_k(1 - z_\ell) + (1 - z_k) z_\ell\} \frac{(\rho - F(\tau)) \mathbb{1}_{\{\rho > F(\tau)\}}}{1 - F(\tau)} \\
c_\gamma(y_k, y_\ell) &= d_\gamma(y_k, y_\ell) - m'_\gamma(y_k, y_\ell) m'_\gamma(y_\ell, y_k) = o_\gamma(1),
\end{aligned}
$$

and A3.3 is verified.

Cut-off sampling for $y_k \leqslant \tau$ is essentially the complement of stratified sampling with a "take-all stratum": $I_{\gamma k} = 1$ for the set $\{k \in U_\gamma : z_k = 1\}$. Take-all strata are common in practice, particularly for the highly-skewed populations in which cut-off sampling is attractive. Arguments nearly identical to those above can be used to establish A3.3 in the take-all case. This take-all stratified design is analogous to the well-known class of *case-control studies* in epidemiology. We specifically consider prospective case-control studies (e.g. Arratia et al. (2005); Langholz and Goldstein (2001); Mantel (1973)), in which the finite population of all disease cases and controls is stratified, disease cases ($z_k = 1$) are selected with probability one, and controls ($z_k = 0$) are selected with probability less than one.

3.2.5 With-replacement sampling with probability proportional to size

Let $\{n^*_\gamma\}$ be a non-random sequence of positive integers with $n^*_\gamma < N_\gamma$ and suppose that f has strictly positive support: $\int \mathbb{1}_{(-\infty,0]} f \, \mathrm{d}\lambda = 0$. Consider the case of with-replacement sampling of n^*_γ draws, with probability of selecting element k on the hth draw equal $p_{\gamma k} \in [0, 1]$, $\sum_{k \in U_\gamma} p_{\gamma k} = 1$. While $p_{\gamma k}$ could be constructed in many ways, a case of particular interest is $p_{\gamma k} \propto Y_k$. This design is usually not feasible in practice, but statistical agencies often attempt to draw samples with probability proportional to a size measure (pps) that is highly correlated with Y. Such a design will be highly efficient for estimation of the Y-total (indeed, a fixed-size pps design with probabilities proportional to Y_k would exactly reproduce the Y-total).

Assume $\mathcal{Z}_\gamma = \mathcal{Y}_\gamma$ and for $z \in (\mathbb{R}^+)^{N_\gamma}$, $D_\gamma(z) = \mathrm{SWR}_{z, n^*_\gamma}$, where for $z \in \mathbb{R}^N$, $n^* \in \mathbb{N}$, $i \in \mathbb{N}^N$,

$$
\mathrm{SWR}_{z,n^*} : \{i\} \mapsto \left(\mathbb{1}_{\{n^*\}}\left(\sum_{k=1}^N i_k\right)\right) \binom{n^*}{i} \prod_{k=1}^N \left(\frac{z_k}{\sum_{k=1}^N z_k}\right)^{i_k}
$$

with the convention $0^0 = 1$ and $\binom{n^*}{i} = n^*! \left(\prod_{k=1}^N i_k!\right)^{-1}$.

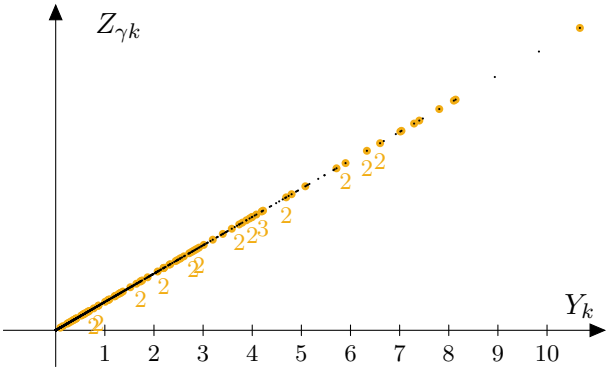
Define $W_{\gamma A} = N_\gamma^{-1} \sum_{j \in U_\gamma : j \notin A} Y_j$. Then

$$
\begin{aligned}
m_\gamma(y_k) &= \frac{n_\gamma^*}{N_\gamma} y_k \, \mathrm{E}\left[\frac{1}{N_\gamma^{-1} y_k + W_{\gamma\{k\}}} \right] \\
m'_\gamma(y_k, y_\ell) &= \frac{n_\gamma^*}{N_\gamma} y_k \, \mathrm{E}\left[\frac{1}{N_\gamma^{-1}(y_k + y_\ell) + W_{\gamma\{k,\ell\}}} \right] \\
v_\gamma(y_k) &= \left(\frac{n_\gamma^*}{N_\gamma} y_k \right)^2 \mathrm{Var}\left[\frac{1}{N_\gamma^{-1} y_k + W_{\gamma\{k\}}} \right] + \frac{n_\gamma^*}{N_\gamma} \frac{y_k}{N_\gamma} \mathrm{E}\left[\frac{W_{\gamma\{k\}}}{\left(N_\gamma^{-1} y_k + W_{\gamma\{k\}} \right)^2} \right] \\
c_\gamma(y_k, y_\ell) &= \left(\frac{n_\gamma^*}{N_\gamma} \right)^2 y_k y_\ell \left\{ n_\gamma^* \mathrm{Var}\left[\frac{1}{N_\gamma^{-1}(y_k + y_\ell) + W_{\gamma\{k,\ell\}}} \right] \right. \\
&\qquad \left. - \frac{1}{N_\gamma} \mathrm{E}\left[\frac{1}{\left(N_\gamma^{-1}(y_k + y_\ell) + W_{\gamma\{k,\ell\}} \right)^2} \right] \right\}.
\end{aligned}
$$

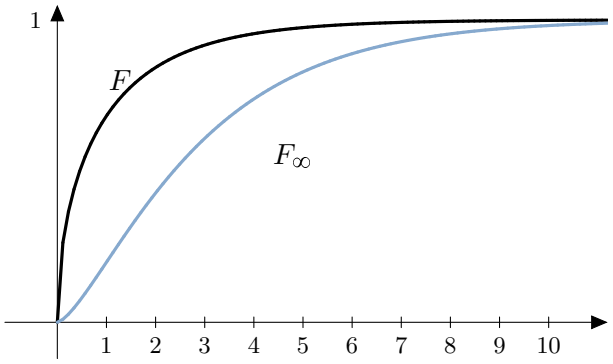
Under mild additional conditions, A3.1 and A3.2 can be established using straightforward bounding and limiting arguments. A sufficient set of conditions for either A3.1 or A3.2 is $n_\gamma^* N_\gamma^{-1} \to \tau \in [0, 1]$ as $\gamma \to \infty$ and $\mathrm{E}\left[Y_1^6 \right] < \infty$. Under these conditions, $m_\gamma(y) = \tau y (\mathrm{E}\left[Y_1 \right])^{-1} + o_\gamma(1)$, and the limiting cdf is the same as in length-biased sampling, as given by equation (3.12). For details, see section C.3.1, appendix C.

Assume for example that $Y_{\gamma 1} \sim \chi^2(1)$. Then, A3.1 and A3.2 are verified. In that case, it is possible to show that $\rho_\infty(y) = y$ and $F_\infty(y) = \pi^{-1/2} \int_0^y t^{1/2}\, e^{-t}\, \mathrm{d}t$, which is the cdf of $\mathrm{Gamma}(3/2, 2)$ distribution. In Figure 3.4(c), the response and study variables on the population and on the sample are plotted. In Figures 3.3(b)–3.4(d), F_γ, F, F_∞ and the population empirical cdf ($\mathbb{R} \to [0,1], \alpha \mapsto N_\gamma^{-1} \sum_{k \in U_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_k)$) are plotted for large values of N_γ (1000) and n_γ (100).

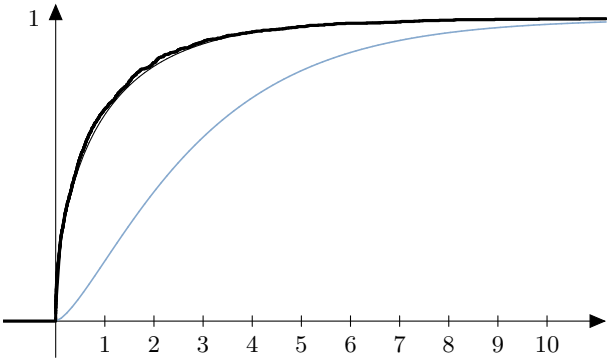
Figure 3.2: With-replacement sampling with probability proportional to size
$Y_\gamma \sim \chi^2$, $Z_{\gamma k} = Y_k$, $N_\gamma = 1000$, $n_\gamma = 100$.



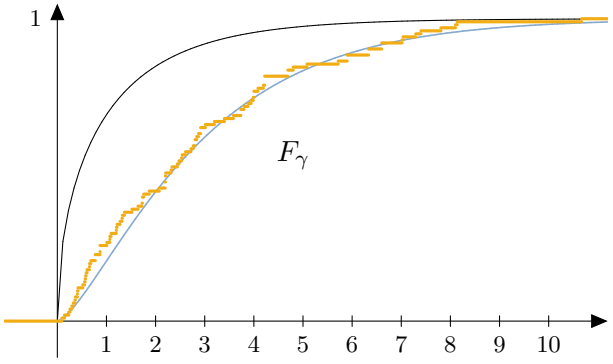
(a) Plot of $(Y_k, Z_{\gamma k})_{k \in U_\gamma}$. The large circles correspond to units selected more than once, the numbers below indicate the number of times units that were selected.



(b) F and F_∞.



(c) Plot of population empirical cdf, F and F_∞.

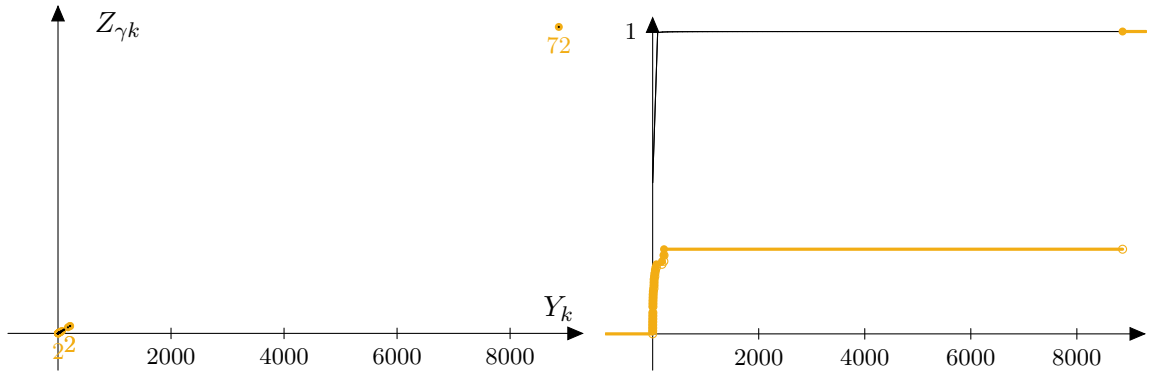


(d) Plot of sample empirical cdf F_γ, F and F_∞.

Assume now that the distribution of Y is the distribution of the absolute value of a random variable that follows a Cauchy distribution. Then A3.0 is not verified, F_∞ is not defined, and the empirical sample cdf does not converge to a limit sample cdf.

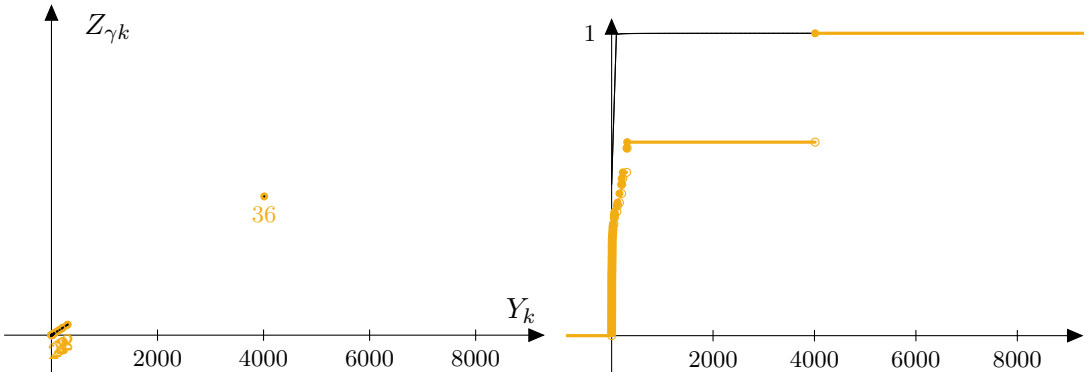
To illustrate that result, the response and design variables for two populations of size $N_\gamma = 1000$ are simulated according to this superpopulation model. For each population, a sample of size $n_\gamma = 100$ is drawn, according to the design measure $\Pi = D(\mathcal{Z}_\gamma)$. In Figure 3.3, the response and design variables on the population and on the sample, F_γ, and F are plotted. In both cases, we notice that there exists an element $k_0 \in U_\gamma$ such that $Y_{\gamma k_0}$ is large with respect to $\sum_{k \in U_\gamma \setminus \{k_0\}} Y_{\gamma k}$, so that k_0 will be selected many times. This induces a high variability of the population responses and the random design measure, and consequently a high variability of the empirical sample cdf.

Figure 3.3: With-replacement sampling with probability proportional to size and Cauchy distribution $\mathrm{P}^{Y_k}(]-\infty, y]) = 2\pi^{-1}\arctan(2y)\mathbb{1}_{[0,+\infty[}(y)$, $Z_{\gamma k} = Y_k$, $N = 1000$, $n = 100$.



(a) First draw. Plot of $(Y_k, Z_{\gamma k})_{k \in U_\gamma}$. The large circles correspond to units selected more than once, the numbers below indicate the number of times units were selected.

(b) First draw. Plot F_γ and F.



(c) Second draw. Plot of $(Y_k, Z_{\gamma k})_{k \in U_\gamma}$. The large circles correspond to units selected more than once, the numbers below indicate the number of times units were selected.

(d) Second draw. Plot F_γ, and F.

3.2.6 Endogenous stratification

Endogenous stratification, in which the sample is effectively stratified on the value of the dependent variable, is common in the health and social sciences (e.g., Hausman and Wise, 1981; Shaw, 1988). Often, it arises by design when a screening sample is selected, the dependent variable is observed, and then covariates are measured for a sub-sample that is stratified on the dependent variable: for example, undersampling the

high-income stratum (Hausman and Wise, 1981). It can also arise through uncontrolled selection effects, in much the same way as length-biased sampling. One such example comes from fisheries surveys, in which a field interviewer is stationed at a dock for a fixed length of time, and intercepts recreational fishing boats as they return to the dock. The interviewer tends to select high-catch boats and, while busy measuring the fish caught on those boats, misses more of the low-catch boats. Thus, sampling effort is endogenously stratified on catch (Sullivan et al., 2006).

We consider a sample endogenously stratified on the order statistics of Y. Let $\{H_\gamma\}$ be a non-random sequence of positive integers, which may remain bounded or go to infinity. For each γ, let $\{N_{\gamma h}\}_{h=1}^{H_\gamma}$ be a set of non-random positive integers with $\sum_{h=1}^{H_\gamma} N_{\gamma h} = N_\gamma$, and let $\{n_{\gamma h}\}_{h=1}^{H_\gamma}$ be a set of non-random positive integers with $n_{\gamma h} \leqslant N_{\gamma h}$. Let

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(N_\gamma)}$$

denote the order statistics for the γth population, which is stratified by taking the first $N_{\gamma 1}$ values as stratum 1, the next $N_{\gamma 2}$ as stratum 2, etc., with the last $N_{\gamma H_\gamma}$ values constituting stratum H_γ. The γth sample is then a stratified simple random sample without replacement of size $n_{\gamma h}$ from the $N_{\gamma h}$ elements in stratum h.

Define $M_{\gamma 0} = 0$ and $M_{\gamma h} = \sum_{g=1}^{h} N_{\gamma h}$. Because H_γ, N_γ and n_γ are not random, we then have

$$\begin{aligned} m_\gamma(y) &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathrm{P}\left(Y_{(M_{\gamma,h-1})} < Y_k \leqslant Y_{(M_{\gamma h})} \mid Y_k = y\right) \\ &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathrm{P}\left(\frac{M_{\gamma,h-1}}{N_\gamma - 1} < F_{N_\gamma - 1}(y) \leqslant \frac{M_{\gamma,h}}{N_\gamma - 1}\right), \end{aligned}$$

where $F_{N_\gamma - 1}(\cdot)$ is the empirical cumulative distribution function for $\{Y_j\}_{j \in U_\gamma : j \neq k}$. From the classical Glivenko-Cantelli theorem, $F_{N_\gamma - 1}(y)$ converges uniformly almost surely to F. Similar computations can be used to derive $m'_\gamma(y_1, y_2)$ and $c_\gamma(y_1, y_2)$ and their limits. With such derivations, it is possible to establish the following result.

Result 3.1. *If* $G(\alpha) = \lim_{\gamma \to \infty} \sum_{h=1}^{H_\gamma} n_{\gamma h} N_{\gamma h}^{-1} \mathbb{1}_{\left(N_\gamma^{-1} M_{\gamma,h-1}, N_\gamma^{-1} M_{\gamma h}\right)}(\alpha)$ *exists except for a finite number of points and is a piecewise continuous non-null function, and the convergence is uniform in* α *then A3.3 and A3.2 hold.*

Proof. See appendix C.3.2. □

Remark

- The assumptions do not specify whether $\lim_\gamma H_\gamma = \infty$ or not. The limit τ may exist in both cases.
- Usually, strata are not based on the ordering of $\mathcal{Y}_\gamma$, but on the ordering of a design variable $\mathcal{Z}_\gamma$ supposed to be correlated to $\mathcal{Y}_\gamma$. We have described two different cases: when $\mathcal{Y}_\gamma$ is ordered like $\mathcal{Z}_\gamma$, and when $\mathcal{Y}_\gamma$ and $\mathcal{Z}_\gamma$ are independent. We can thus expect that in realistic cases, A3.1 will be satisfied when appropriate assumptions on $(n_{\gamma h})$ and $(N_{\gamma h})$ hold.

  It is also worth noting that our assumptions are not necessary. For example, we can construct a case in which assumptions on $(n_{\gamma h})$ and $(N_{\gamma h})$ hold and the dependence between $\mathcal{Y}_\gamma$ and $\mathcal{Z}_\gamma$ is chaotic: we choose $\tau = \mathrm{Id}_{\mathbb{R}}|_{[0,1]}$, $Y_k \sim \mathscr{U}_{[0,1]}$, $\mathcal{Z}_\gamma = 10^{-2\gamma}\left\lfloor 10^{2\gamma}\mathcal{Y}_\gamma - 10^\gamma \left\lfloor 10^\gamma \mathcal{Y}_\gamma \right\rfloor \right\rfloor$, $N_\gamma = \gamma^2$, $H_\gamma = \gamma$, $N_{\gamma h} = \gamma$, $n_{\gamma h} = h$. In this case, we do not have a pointwise convergence of m_γ, e.g. A3.0 fails. Nevertheless, even in that case, we can show that A2.0 holds and that $P^{Y_1|I_{\gamma 1}=1} \xrightarrow{weakly} P^{Y_1}$, by showing that for all interval A, $\lim_\gamma P^{Y_1|I_{\gamma 1}=1}(A) = P^{Y_1}(A)$. We can also show the uniform L^2 and almost sure convergence of the sample cdf to F. Therefore, our conditions are sufficient but not necessary.

3.3 Conclusion

Assumptions on the selection mechanism and the superpopulation model under which the unweighted empirical sample cdf converges uniformly to a weighted version of the superpopulation cdf have been given. Because the conditions that have been specified on the informative selection mechanism are closely tied to first and second-order inclusion probabilities in a standard design-based survey sampling setting, the conditions are verifiable. The given examples illustrate the computations for selection mechanisms encountered in real surveys and observational studies. The versions of Glivenko-Cantelli theorem adapted to non-informative selection under the superpopulation and the fixed population models are a first indication that the response values may behave asymptotically as if they were iid $(\rho_\infty f.\lambda)$.

References

Arratia, R., Goldstein, L., and Langholz, B. (2005). Local central limit theorems, the high-order correlations of rejective sampling and logistic likelihood asymptotics. *The Annals of Statistics*, 33(2):871–914.

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28(4):1026–1053.

Breidt, F. J. and Opsomer, J. D. (2008). Endogenous post-stratification in surveys: classifying with a sample-fitted model. *The Annals of Statistics*, 36(1):403–427.

Cox, D. (1969). *Some Sampling Problems in Technology*, pages 506–527. Wiley Interscience.

Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons Inc.

Hájek, J. (1981). *Sampling from a Finite Population*, volume 37 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Edited by Václav Dupač, With a foreword by P. K. Sen.

Hausman, J. and Wise, D. (1981). *Stratification on endogenous variables and estimation: The Gary income maintenance experiment*, pages 36–391. Cambridge, MA: MIT Press.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77(377):89–96.

Langholz, B. and Goldstein, L. (2001). Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics*, 2(1):63–84.

Leigh, G. M. (1988). A comparison of estimates of natural mortality from fish tagging experiments. *Biometrika*, 75(2):347–353.

Mantel, N. (1973). Synthetic retrospective studies and related topics. *Biometrics*, 29:479–486.

Nowell, C. and Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research*, 28(4):475–479.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189.

Robinson, P. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā: The Indian Journal of Statistics, Series B*, 45(2):240–248.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer-Verlag, New York.

Shaw, D. (1988). On-site samples' regression: problems of nonnegative integers, truncation, and endogenous stratification. *Journal of Econometrics*, 37(2):211–223.

Sullivan, P., Breidt, F. J., Ditton, R., Knuth, B., Leaman, B., O'Connell, V., Parsons, G., Pollock, K., Smith, S., and S.Stokes (2006). *Review of Recreational Fisheries Survey Methods*. National Academies Press, Washington, DC.

Chapter 4

Kernel density estimation

In this chapter, we consider the same superpopulation model as defined in chapter 3, section 3.1.1. The functions ρ_γ, m_γ, m'_γ, c_γ, d_γ, c_γ, and their limit versions are also defined as in section 3.1.1. We state some conditions on the sequence of sample schemes under which the sample kernel density estimator converges locally in L_2 to the sample pdf. We adapt the Bochner Lemma (see Theorems 4.1, 4.2) to take into account the non-independence of the observations on the sample, we prove the asymptotic unbiasedness of the sample kernel density estimator of the limit sample pdf, and we give the rate of convergence of the sample kernel density estimator in the case of informative selection to the limit sample pdf. The results are an adaptation of Tsybakov (2009, chapter 1). We illustrate our results by showing that the conditions stated can be easily verified or rejected for some specified designs. For these specified designs, we compare the approximation of variance we obtain to a computation of the variance by Monte Carlo methods.

4.1 Definitions

Definition 4.1. *Let T be an interval of $\mathbb{R}$, $\beta \in]0, +\infty[$, $L \in]0, +\infty[$. The Hölder class on T, denoted $\Sigma(\beta, L, T)$ is the set of functions $g : T \to \mathbb{R}$ such that $g^{(l)}$ exists (with $l = \lfloor \beta \rfloor$, denoting the largest integer less than or equal to β) and $\forall (x, x') \in T^2$, $\left| g^{(l)}(x) - g^{(l)}(x') \right| \leqslant L \left| x - x' \right|^{\beta - l}$. Let $\mathcal{P}(\beta, L) = \left\{ g \in \Sigma(\beta, L, \mathbb{R}) | g \geqslant 0, \int g d\lambda = 1 \right\}$.*

Definition 4.2. *A kernel K is a measurable function $K : \mathbb{R} \to \mathbb{R}$ such that $\int K \, \mathrm{d}\lambda = 1$. For $l \in \mathbb{N}$, K is a kernel of order l if $\int K \, \mathrm{d}\lambda = 1$, and $\forall j \in \{1, \ldots, l\}$, $\int u^j K(u) \, \mathrm{d}\lambda(u) = 0$.*

Definition 4.3. *A sequence of bandwidths is a sequence $(h_\gamma)_{\gamma \in \mathbb{N}} \in]0, +\infty[^{\mathbb{N}}$.*

In the following K is a kernel and $(h_\gamma)_{\gamma \in \mathbb{N}}$ is a sequence of bandwidths such that $\lim_{\gamma \to \infty} h_\gamma = 0$.

Definition 4.4. *The kernel density estimator of $\rho_\gamma f$ associated to K and $(h_\gamma)_{\gamma \in \mathbb{N}}$ is*

$$p_\gamma : y \mapsto \frac{1}{h_\gamma \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma) \right)} \sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y}{h_\gamma} \right) I_{\gamma k}.$$

4.2 Properties of the kernel estimator

Proposition 4.1. $\forall y_0 \in \mathbb{R}$,

$$\begin{aligned}\operatorname{Var}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] \;\leqslant\; & \frac{\sup_{y\in\mathbb{R}}\{(v_\gamma \times f)(y)\} + \sup_{y\in\mathbb{R}}\{(m_\gamma \times f)(y)\}\int K^2 d\lambda}{N_\gamma h_\gamma \left(\int m_\gamma f d\lambda\right)^2} \\ & + \frac{1}{\left(\int m_\gamma f d\lambda\right)^2} \sup_{(y_1,y_2)\in\mathbb{R}^2}\{(c_\gamma(y_1,y_2)f(y_1)f(y_2))\} \\ & + \frac{1}{\left(\int m_\gamma f d\lambda\right)^2} \sup_{(y_1,y_2)\in\mathbb{R}^2}\{\left(m'_\gamma(y_2,y_1) - m_\gamma(y_1)m_\gamma(y_2)\right) f(y_1)f(y_2))\}.\end{aligned}$$

Proof.

$$\begin{aligned}&\operatorname{Var}\left[\frac{n_\gamma}{N_\gamma\int m_\gamma f d\lambda}p_\gamma(y_0)\right] \\ &= \operatorname{Var}\left[\frac{1}{h_\gamma N_\gamma \int m_\gamma f d\lambda}\sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) I_{\gamma k}\right] \\ &= \frac{1}{N_\gamma\left(h_\gamma\int m_\gamma f d\lambda\right)^2}\operatorname{Var}\left[K\left(\frac{Y_1 - y_0}{h_\gamma}\right) I_{\gamma 1}\right] \\ &\quad + \frac{N_\gamma - 1}{N_\gamma\left(h_\gamma\int m_\gamma f d\lambda\right)^2} Cov\left[K\left(\frac{Y_1 - y_0}{h_\gamma}\right) I_{\gamma 1}, K\left(\frac{Y_2 - y_0}{h_\gamma}\right) I_{\gamma 2}\right] \\ &= \frac{1}{N_\gamma\left(h_\gamma\int m_\gamma f d\lambda\right)^2}\left(\int K\left(\frac{y - y_0}{h_\gamma}\right) v_\gamma(y) f(y) d\lambda(y)\right. \\ &\qquad + \int K^2\left(\frac{y - y_0}{h_\gamma}\right) m_\gamma^2(y) f(y) d\lambda(y) \\ &\qquad \left. - \left(\int K\left(\frac{y - y_0}{h_\gamma}\right) m_\gamma(y) f(y) d\lambda(y)\right)^2\right) \\ &\quad + \frac{N_\gamma - 1}{N_\gamma\left(h_\gamma\int m_\gamma f d\lambda\right)^2}\left(\int K\left(\frac{y_1 - y_0}{h_\gamma}\right) K\left(\frac{y_2 - y_0}{h_\gamma}\right) c_\gamma(y_1,y_2) f(y_1) f(y_2) d\lambda(y_1) d\lambda(y_2)\right. \\ &\qquad + \int K\left(\frac{y_1 - y_0}{h_\gamma}\right) K\left(\frac{y_2 - y_0}{h_\gamma}\right) \left(m'_\gamma(y_1,y_2) m'_\gamma(y_2,y_1)\right. \\ &\qquad \left.\left. - m_\gamma(y_1) m_\gamma(y_2)\right) f(y_1) f(y_2) d\lambda(y_1) d\lambda(y_2)\right).\end{aligned}$$

□

Proposition 4.2. *Assume that K is a kernel of order l such that $\int |u|^\beta |K(u)|\, du < +\infty$. If $\rho_\gamma f \in \mathcal{P}(\beta, L)$, then*

$$\forall y_0 \in \mathbb{R},\ \forall \gamma \in \mathbb{N},\ \left|\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] - \rho_\gamma(y_0) f(y_0)\right| \leqslant C_2 h_\gamma^\beta,$$

with $C_2 = \frac{L}{l!}\int |u|^\beta |K(u)| du$.

Proof.

$$\begin{aligned}
&\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\\
&= \mathrm{E}\left[\frac{n_\gamma}{N_\gamma \int m_\gamma f d\lambda}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\\
&= \frac{1}{h_\gamma}\int K\left(\frac{y-y_0}{h_\gamma}\right)(\rho_\gamma \times f)(y)\,\mathrm{d}y - (\rho_\gamma \times f)(y_0)\\
&= \int K(u)\left((\rho_\gamma \times f)(y_0+uh_\gamma) - (\rho_\gamma \times f)(y_0)\right)\mathrm{d}u.
\end{aligned}$$

In addition, $\exists \tau_\gamma : u \mapsto [0,1]$ such that:

$$(\rho_\gamma \times f)(y_0+uh_\gamma) = \sum_{j=0}^{l-1}\frac{(\rho_\gamma \times f)^{(j)}(y_0)(uh_\gamma)^j}{j!} + \frac{(\rho_\gamma \times f)^{(l)}(y_0+\tau_\gamma(u)uh_\gamma)(uh_\gamma)^l}{l!}.$$

As K is of order $l = \lfloor\beta\rfloor$,

$$\begin{aligned}
&\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\\
&= \int K(u)\frac{(uh_\gamma)^l}{l!}\left((\rho_\gamma \times f)^{(l)}(y_0+\tau_\gamma(u)uh_\gamma) - (\rho_\gamma \times f)^{(l)}(y_0)\right)\mathrm{d}u,
\end{aligned}$$

and

$$\begin{aligned}
&\left|\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\right|\\
&\leqslant \int |K(u)|\frac{|uh|^l}{l!}\left|(\rho_\gamma \times f)^{(l)}(y_0+\tau_\gamma(u)uh_\gamma) - (\rho_\gamma \times f)^{(l)}(y_0)\right|\mathrm{d}u\\
&\leqslant L\int |K(u)|\frac{|uh|^l}{l!}\left|\tau_\gamma(u)uh_\gamma\right|^{\beta-l}\mathrm{d}u\\
&\leqslant C_2 h_\gamma^\beta.
\end{aligned}$$

□

Proposition 4.3. *If $f \in \mathcal{P}(\beta, L)$, then*

$$\begin{aligned}
&\sup_{y_0\in\mathbb{R}}\mathrm{E}\left[\left(\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0) - (\rho_\gamma \times f)(y_0)\right)^2\right]\\
&\leqslant \frac{\sup_{y\in\mathbb{R}}\{(v_\gamma \times f)(y)\} + \sup_{y\in\mathbb{R}}\{(m_\gamma \times f)(y)\}\int K^2\,\mathrm{d}\lambda}{N_\gamma h_\gamma\left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2}\\
&\quad + \frac{1}{\left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2}\sup_{(y_1,y_2)\in\mathbb{R}^2}\{(c_\gamma(y_1,y_2)f(y_1)f(y_2))\}\\
&\quad + \frac{1}{\left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2}\sup_{(y_1,y_2)\in\mathbb{R}^2}\left\{\left(m'_\gamma(y_2,y_1) - m_\gamma(y_1)m_\gamma(y_2)\right)f(y_1)f(y_2))\right\}\\
&\quad + C_2^2 h_\gamma^{2\beta}.
\end{aligned}$$

Proof. Let $y_0 \in \mathbb{R}$, $\gamma \in \mathbb{N}$. Then by Propositions 4.1 and 4.2,

$$\begin{aligned}
&\mathrm{E}\left[\left(\frac{n_\gamma}{\mathrm{E}[n_\gamma]}p_\gamma(y_0) - (\rho_\gamma \times f)(y_0)\right)^2\right] \\
&\leqslant \quad \mathrm{Var}\left[\frac{n_\gamma}{\mathrm{E}[n_\gamma]}p_\gamma(y_0)\right] + \left|\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}[n_\gamma]}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\right|^2 \\
&\leqslant \quad \frac{\sup_{y\in\mathbb{R}}\{(v_\gamma \times f)(y)\} + \sup_{y\in\mathbb{R}}\{(m_\gamma \times f)(y)\}\int K^2\,\mathrm{d}\lambda}{N_\gamma h_\gamma \left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2} \\
&\qquad + \frac{1}{\left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2}\sup_{(y_1,y_2)\in\mathbb{R}^2}\{(c_\gamma(y_1,y_2)f(y_1)f(y_2))\} \\
&\qquad + \frac{1}{\left(\int m_\gamma f\,\mathrm{d}\lambda\right)^2}\sup_{(y_1,y_2)\in\mathbb{R}^2}\left\{\left(m'_\gamma(y_2,y_1) - m_\gamma(y_1)m_\gamma(y_2)\right)f(y_1)f(y_2))\right\} \\
&\qquad + C_2^2 h_\gamma^{2\beta}.
\end{aligned}$$

□

Theorem 4.1. *Bochner lemma in the case of informative sampling.*

Let $y_0 \in \mathbb{R}$. Assume that g is a real function, continuous on a neighborhood of $y_0 \in \mathbb{R}$, $Q : \mathbb{R} \to \mathbb{R}$ is measurable, $\int |Q|\,\mathrm{d}\lambda < +\infty$, r_γ is a sequence of real measurable functions, and r is a real measurable function such that:

$$\begin{cases}
r_\gamma - r \text{ converges uniformly on a neighborhood of } y_0 \text{ to } 0, \\
r \text{ is continuous in } y_0, \\
\sup\limits_{u\in\mathbb{R}}(g \times r_\gamma)(u) = O_\gamma(1), \\
\sup\limits_{u\in\mathbb{R}}(g \times r) < +\infty.
\end{cases}$$

Then:

$$\lim_{\gamma\to\infty}\frac{1}{h_\gamma}\int Q\left(\frac{y-y_0}{h_\gamma}\right)r_\gamma(y)g(y)\,\mathrm{d}\lambda(y) = r(y_0)g(y_0)\int Q(y)\,\mathrm{d}\lambda(y) + o_\gamma(1).$$

Proof. We calculate:

$$\begin{aligned}
&\left|\frac{1}{h_\gamma}\int Q\left(\frac{y-y_0}{h_\gamma}\right)r_\gamma(y)f(y)\,\mathrm{d}\lambda(y) - g(y_0)r_\gamma(y_0)\int Q(u)\,\mathrm{d}\lambda(u)\right| \\
&= \quad \left|\int \left((g \times r_\gamma)(y_0 + uh_\gamma) - (g \times r)(y_0)\right)Q(u)\,\mathrm{d}\lambda(u)\right| \\
&\leqslant \quad \sup_{|u|\leqslant h_\gamma^{-1/2}}\{|(g \times r_\gamma)(y_0 + uh_\gamma) - (g \times r)(y_0)|\}\int |Q|(u)\,\mathrm{d}\lambda(u) \\
&\qquad + \int_{|u|>h_\gamma^{-1/2}} |(g \times r_\gamma)(y_0 + uh_\gamma) - (g \times r)(y_0)|\,|Q|(u)\,\mathrm{d}\lambda(u) \\
&\leqslant \quad \sup_{|v|\leqslant h_\gamma^{1/2}}\{|(g \times r_\gamma)(y_0 + v) - (g \times r)(y_0)|\}\int |Q|(u)\,\mathrm{d}\lambda(u) \\
&\qquad + \left(\sup_{u\in\mathbb{R}}\{(g \times r_\gamma)(u)\} + \sup_{u\in\mathbb{R}}\{(g \times r)(u)\}\right)\int_{|u|>h_\gamma^{-1/2}} |Q|(u)\,\mathrm{d}\lambda(u)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \left(\sup_{|v|\leqslant h_\gamma^{1/2}} \{|(g\times r_\gamma)(y_0+v)-(g\times r)(y_0+v)|\} \right.\\
&\qquad \left. + \sup_{|v|\leqslant h_\gamma^{1/2}} \{|(g\times r)(y_0+v)-(g\times r)(y_0)|\} \right) \int |Q|(u)\,\mathrm{d}\lambda(u)\\
&\quad + \left(\sup_{u\in\mathbb{R}} \{(g\times r_\gamma)(u)\} + \sup_{u\in\mathbb{R}} \{(g\times r)(u)\} \right) \int_{|u|>h_\gamma^{-1/2}} |Q|(u)\,\mathrm{d}\lambda(u)\\
&\leqslant \left(\sup_{|v|\leqslant h_\gamma^{1/2}} \{|(r_\gamma - r)(y_0+v)|\} \sup_{|v|\leqslant h_\gamma^{1/2}} \{|g(y_0+v)|\} \right.\\
&\qquad \left. + \sup_{|v|\leqslant h_\gamma^{1/2}} \{|(g\times r)(y_0+v)-(g\times r)(y_0)|\} \right) \int |Q|(u)\,\mathrm{d}\lambda(u)\\
&\quad + \left(\sup_{u\in\mathbb{R}} \{(g\times r_\gamma)(u)\} + \sup_{u\in\mathbb{R}} \{(g\times r)(u)\} \right) \int_{|u|>h_\gamma^{-1/2}} |Q|(u)\,\mathrm{d}\lambda(u).
\end{aligned}
$$

When $h_\gamma \to 0$ every summand tends to 0. □

Theorem 4.2. *Let $y_0 \in \mathbb{R}$, Assume that g is a real function, continuous on a neighborhood of $y_0 \in \mathbb{R}$, $Q : \mathbb{R} \to \mathbb{R}$ is measurable, $\int |Q|\,\mathrm{d}\lambda < +\infty$, $r_\gamma : \mathbb{R}^2 \to \mathbb{R}$ is a sequence of measurable functions and:*

$$
\begin{cases}
r_\gamma \text{ converges uniformly on a neighborhood of } (y_0,y_0) \text{ to } 0,\\
\sup\limits_{(u_1,u_2)\in\mathbb{R}^2} \{|r_\gamma(u_1,u_2)g(u_1)g(u_2)|\} = O_\gamma(1).
\end{cases}
$$

Then:

$$
\frac{1}{h_\gamma^2} \int Q\left(\frac{y_1-y_0}{h_\gamma}\right) Q\left(\frac{y_2-y_0}{h_\gamma}\right) r_\gamma(y_1,y_2)g(y_1)g(y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 = o_\gamma(1).
$$

Proof. We calculate:

$$
\begin{aligned}
&\left| \frac{1}{h_\gamma^2} \int Q\left(\frac{y_1-y_0}{h_\gamma}\right) Q\left(\frac{y_2-y_0}{h_\gamma}\right) r_\gamma(y_1,y_2)g(y_1)g(y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \right|\\
&\quad = \left| \int Q(y_0+u_1h_\gamma)\,Q(y_0+u_2h_\gamma)\,r_\gamma(y_0+u_1h_\gamma,y_0+u_2h_\gamma)g(y_0+u_1h_\gamma)g(y_0+u_2h_\gamma)\,\mathrm{d}u_1\,\mathrm{d}u_2 \right|\\
&\quad \leqslant \sup_{\max(u_1,u_2)\leqslant h_\gamma^{-1/2}} \{|g(y_0+u_1h_\gamma)g(y_0+u_2h_\gamma)r_\gamma(y_0+u_1h_\gamma,y_0+u_2h_\gamma)|\} \left(\int |Q|(u)\,\mathrm{d}u \right)^2\\
&\qquad + \sup_{(u_1,u_2)\in\mathbb{R}^2} \{|g(u_1)g(u_2)r_\gamma(u_1,u_2)|\} \left(\int_{u>h_\gamma^{-1/2}} |Q|(u)\,\mathrm{d}u \right)^2\\
&\quad \leqslant \sup_{\max(v_1,v_2)\leqslant h_\gamma^{1/2}} \{|g(y_0+v_1)g(y_0+v_2)r_\gamma(y_0+v_1,y_0+v_2)|\} \left(\int |Q|(u)\,\mathrm{d}u \right)^2\\
&\qquad + \sup_{(u_1,u_2)\in\mathbb{R}^2} \{|g(u_1)g(u_2)r_\gamma(u_1,u_2)|\} \left(\int_{u>h_\gamma^{-1/2}} |Q|(u)\,\mathrm{d}u \right)^2.
\end{aligned}
$$

When $h_\gamma \to 0$ every summand tends to 0. □

Proposition 4.4. *Limit of the expected value of the sample kernel density estimator*
Let $y_0 \in \mathbb{R}$. We assume A3.0 and:

$$\bullet\,\mathbf{A4.3}\begin{cases} \operatorname{Var}[n_\gamma] = o_\gamma(N_\gamma^2), & (4.3a)\\ \sup_{y\in\mathbb{R}}\{(m_\gamma f)(y)\} = O_\gamma(1), & (4.3b)\\ \sup_{y\in\mathbb{R}}\{(m_\infty f)(y)\} < +\infty, & (4.3c)\\ m_\gamma - m_\infty \text{ converges uniformly on a neighborhood of } y_0 \text{ to } 0, & (4.3d)\\ m_\infty \text{ is continuous in } y_0, & (4.3e)\\ f \text{ is continuous in } y_0. & (4.3f)\end{cases}$$

Then

$$\lim_{\gamma\to\infty} \mathrm{E}[p_\gamma(y_0)] = \rho_\infty(y_0) f(y_0).$$

Proof. With A3.0, as $\lim_{\gamma\to\infty}\int m_\gamma f > 0$, then $\exists\Gamma\in\mathbb{N}$ such that $\forall\gamma\in\mathbb{N}$, $\gamma\geqslant\Gamma \Rightarrow \mathrm{E}[n_\gamma] > 0$. Then,

$$\mathrm{E}\left[\frac{n_\gamma}{N_\gamma\int m_\gamma f} p_\gamma(y_0)\right] = \left(\frac{1}{\int m_\gamma f} + o_\gamma(1)\right)\frac{1}{h_\gamma}\int K\left(\frac{y-y_0}{h_\gamma}\right) m_\gamma(y) f(y)\,\mathrm{d}\lambda(y).$$

We apply Theorem 4.1, with $g = f$, $r_\gamma = m_\gamma$, $r = m_\infty$, and obtain that:

$$\lim_{\gamma\to\infty} \mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}[n_\gamma]} p_\gamma(y_0)\right] = \rho(y_0) f(y_0).$$

Besides,

$$p_\gamma(y_0) - \frac{n_\gamma}{\mathrm{E}[n_\gamma]} p_\gamma(y_0) = \left(\frac{\mathrm{E}[n_\gamma]}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} - 1\right)\frac{n_\gamma}{\mathrm{E}[n_\gamma]} p_\gamma(y_0),$$

and with A4.3a, $\mathrm{E}\left[\left(\frac{\mathrm{E}[n_\gamma]}{n_\gamma+\mathbb{1}_{\{0\}}(n_\gamma)} - 1\right)^2\right] = o_\gamma(1)$. □

Proposition 4.5. *Let $y_0 \in \mathbb{R}$ and K be a kernel.*
We assume A3.0, A4.3 and that $\exists V$ a neighborhood of y_0 such that:

$$\bullet\,\mathbf{A4.4}\begin{cases} \int K^2(u)\,\mathrm{d}u < +\infty, & (4.4a)\\ \exists v_\infty \text{ a measurable real function such that } v_\gamma - v_\infty \text{ converges uniformly to } 0 \text{ on } V, & (4.4b)\\ v_\infty \text{ is continuous in } y_0, & (4.4c)\\ \sup_{y\in\mathbb{R}}\{v_\gamma(y) f(y)\} = O_\gamma(1), & (4.4d)\\ \sup_{y\in\mathbb{R}}\{v_\infty f(y)\} < +\infty, & (4.4e)\\ \sup_{(y_1,y_2)\in\mathbb{R}^2}\{(\mathrm{d}_\gamma(y_1,y_2) + c_\gamma(y_1,y_2))\} = O_\gamma(1), & (4.4f)\\ c_\gamma + \mathrm{d}_\gamma \text{ converges uniformly to } 0 \text{ on } V\times V. & (4.4g)\end{cases}$$

Then

$$\operatorname{Var}[p_\gamma(y_0)] = \frac{f(y_0)}{N_\gamma h_\gamma}\left(\left(\frac{v_\infty(y_0)}{\left(\int m_\infty f\,\mathrm{d}\lambda\right)^2} + \rho_\infty^2(y_0)\right)\int K^2(u)\,\mathrm{d}u\right)(1 + o_\gamma(1)).$$

Proof. We calculate:

$$\begin{aligned}
&\operatorname{Var}\left[\frac{1}{h_\gamma N_\gamma \int m_\gamma f \,\mathrm{d}\lambda}\sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) I_{\gamma k}\right] \\
&= \frac{1}{N_\gamma\left(\int m_\gamma f \,\mathrm{d}\lambda\right)^2}\left(\frac{1}{h_\gamma^2}\int K\left(\frac{y-y_0}{h_\gamma}\right) v_\gamma(y) f(y)\,\mathrm{d}\lambda(y)\right. \\
&\qquad + \frac{1}{h_\gamma^2}\int K^2\left(\frac{y-y_0}{h_\gamma}\right) m_\gamma^2(y) f(y)\,\mathrm{d}\lambda(y) \\
&\qquad \left. - \left(\frac{1}{h_\gamma}\int K\left(\frac{y-y_0}{h_\gamma}\right) m_\gamma(y) f(y)\,\mathrm{d}\lambda(y)\right)^2\right) \\
&\quad + \frac{N_\gamma - 1}{N_\gamma\left(h_\gamma \int m_\gamma f \,\mathrm{d}\lambda\right)^2}\left(\int K\left(\frac{y_1-y_0}{h_\gamma}\right) K\left(\frac{y_2-y_0}{h_\gamma}\right)(c_\gamma + \mathrm{d}_\gamma)(y_1,y_2) f(y_1) f(y_2)\,\mathrm{d}\lambda(y_1)\,\mathrm{d}\lambda(y_2)\right).
\end{aligned}$$

By application of Theorem 4.1 to $Q = K$, $r_\gamma = v_\gamma$, and $r = v_\infty$, we obtain that:

$$\frac{1}{h_\gamma}\int K\left(\frac{y-y_0}{h_\gamma}\right) v_\gamma(y) f(y)\,\mathrm{d}\lambda(y) = v_\infty(y_0) f(y_0) + o_\gamma(1).$$

By application of Theorem 4.1 to $Q = K$, $r_\gamma = m_\gamma$, and $m = v$, we obtain that:

$$\frac{1}{h_\gamma}\int K\left(\frac{y-y_0}{h_\gamma}\right) m_\gamma(y) f(y)\,\mathrm{d}\lambda(y) = m(y_0) f(y_0) + o_\gamma(1).$$

By application of Theorem 4.2 to $Q = K$, and $r_\gamma = (c_\gamma + \mathrm{d}_\gamma)$, we obtain that:

$$\frac{1}{(h_\gamma)^2}\left(\int K\left(\frac{y_1-y_0}{h_\gamma}\right) K\left(\frac{y_2-y_0}{h_\gamma}\right)(c_\gamma + \mathrm{d}_\gamma)(y_1,y_2) f(y_1) f(y_2)\,\mathrm{d}\lambda(y_1)\,\mathrm{d}\lambda(y_2)\right) = o_\gamma(1).$$

So

$$\begin{aligned}
&\operatorname{Var}\left[\frac{1}{h_\gamma N_\gamma \int m_\gamma f \,\mathrm{d}\lambda}\sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) I_{\gamma k}\right] \\
&= \frac{1}{N_\gamma h_\gamma\left(\int m_\gamma f \,\mathrm{d}\lambda\right)^2}\left(v(y_0) f(y_0) + o_\gamma(1)\right) \\
&\qquad + \frac{1}{N_\gamma h_\gamma}\left(\rho^2(y_0) f(y_0)\int K^2(u)\,\mathrm{d}u\right)(1 + o_\gamma(1)) \\
&\qquad + \frac{1}{N_\gamma}\left(\rho^2(y_0) f^2(y_0) + o_\gamma(1)\right) \\
&\qquad + o_\gamma(1).
\end{aligned}$$

Thus

$$\operatorname{Var}\left[\frac{n_\gamma}{\mathrm{E}[n_\gamma]} p_\gamma(y)\right] \sim_\gamma \frac{f(y_0)}{N_\gamma h_\gamma}\left(\left(\frac{v(y_0)}{\left(\int m_\infty f \,\mathrm{d}\lambda\right)^2} + \rho_\infty^2(y_0)\right)\int K^2(u)\,\mathrm{d}u\right)$$

and by A4.3a, $\mathrm{E}\left[\left(\frac{n_\gamma}{\mathrm{E}[n_\gamma]} - 1\right)^2\right] = o_\gamma(1)$, which together with the preceding implies the result. □

Proposition 4.6. *Let $y_0 \in \mathbb{R}$. We assume that K is a kernel of order $l = 1$. We assume A3.0 and:*

$$\bullet\mathbf{A4.5}\begin{cases} \mathrm{Var}\left[n_\gamma\right] = o_\gamma(N_\gamma^2), & (4.5a)\\ (\rho_\gamma \times f) \in \mathscr{C}^2, & (4.5b)\\ (\rho_\gamma \times f)^{(2)} - (\rho_\infty \times f)^{(2)} \text{ converges uniformly to } 0 \text{ on a neighborhood of } y_0, & (4.5c)\\ \sup\limits_{u\in\mathbb{R}} \left\{ \left|(\rho_\gamma \times f)^{(2)}(u)\right| \right\} = O_\gamma(1), & (4.5d)\\ \displaystyle\int u^2 K(u)\,\mathrm{d}u < +\infty, & (4.5e)\\ \displaystyle\int u^2 K(u)\,\mathrm{d}u \neq 0, & (4.5f) \end{cases}$$

Where $\mathscr{C}^2$ is the space of twice differentiable and continuous functions. Then

$$\mathrm{E}\left[p_\gamma(y_0) - \rho(y_0)f(y_0)\right] = \frac{h_\gamma^2}{2}\left((\rho_\gamma \times f)^{(2)}(y_0)\int u^2K(u)\,\mathrm{d}u + o_\gamma(1)\right).$$

Proof. We use the calculations of Proposition 4.2 with $l = 2$, and we the fact that K is a kernel of order $l = 1$: $\exists \tau_\gamma : u \mapsto [0,1]$ such that:

$$\begin{aligned} &\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0)\right] - \rho_\gamma(y_0)f(y_0)\\ &= \int K(u)\frac{(uh_\gamma)^2}{2}(\rho_\gamma \times f)^{(2)}(y_0 + \tau(u)uh_\gamma)\,\mathrm{d}u\\ &= \frac{h_\gamma^2}{2}(\rho_\gamma \times f)^{(2)}(y_0)\int u^2K(u)\,\mathrm{d}u + \Delta_\gamma(u), \end{aligned}$$

with

$$\begin{aligned} \Delta_\gamma(u) &= \int u^2K(u)\left((\rho_\gamma \times f)^{(2)}(y_0 + \tau_\gamma(u)uh_\gamma) - (\rho_\gamma \times f)^{(2)}(y_0)\right)\mathrm{d}u\\ |\Delta_\gamma(u)| &\leqslant \sup_{u\leqslant h_\gamma^{-1/2}}\left\{\left|(\rho_\gamma \times f)^{(2)}(y_0 + \tau_\gamma(u)uh_\gamma) - (\rho_\gamma \times f)^{(2)}(y_0)\right|\right\}\int u^2K(u)\,\mathrm{d}u\\ &\quad +2\sup_{u\in\mathbb{R}}\left\{\left|(\rho_\gamma \times f)^{(2)}(u)\right|\right\}\int_{u>h_\gamma^{-1/2}} u^2K(u)\,\mathrm{d}u\\ &\leqslant \sup_{v\leqslant h_\gamma^{1/2}}\left\{\left|(\rho_\gamma \times f)^{(2)}(y_0 + v) - (\rho_\gamma \times f)^{(2)}(y_0)\right|\right\}\int u^2K(u)\,\mathrm{d}u\\ &\quad +2\sup_{u\in\mathbb{R}}\left\{\left|(\rho_\gamma \times f)^{(2)}(u)\right|\right\}\int_{u>h_\gamma^{-1/2}} u^2K(u)\,\mathrm{d}u\\ &= O_\gamma(h_\gamma). \end{aligned}$$

Then

$$\mathrm{E}\left[\frac{n_\gamma}{\mathrm{E}\left[n_\gamma\right]}p_\gamma(y_0) - \rho(y_0)f(y_0)\right] = \frac{h_\gamma^2}{2}\left((\rho_\gamma \times f)^{(2)}(y_0)\int u^2K(u)\,\mathrm{d}u + o_\gamma(1)\right).$$

By A4.3a, $\mathrm{E}\left[\left(\frac{n_\gamma}{\mathrm{E}[n_\gamma]} - 1\right)^2\right] = o_\gamma(1)$, and then:

$$\mathrm{E}\left[p_\gamma(y_0) - \rho(y_0)f(y_0)\right] = \frac{h_\gamma^2}{2}\left((\rho_\gamma \times f)^{(2)}(y_0)\int u^2K(u)\,\mathrm{d}u + o_\gamma(1)\right).$$

□

4.3 Link to the Horvitz Thompson estimator of the sample cdf

When ρ_∞ is known, $p_\gamma(y)/\rho_\infty(y)$ may be compared to a Horvitz-Thompson or Hájek-type estimator of $f(y)$, defined as

$$\hat{f}_\gamma(y_0) = \left(\sum_{k\in U_\gamma} \frac{I_{\gamma k}}{\pi_{\gamma k}}\right)^{-1} \left(\frac{1}{h_\gamma} \sum_{k\in U_\gamma} \frac{I_{\gamma k} K\left(\frac{Y_k - y_0}{h_\gamma}\right)}{\pi_{\gamma k}}\right).$$

When $\rho_\infty(y) = \lim_\gamma \frac{N_\gamma}{n_\gamma} \mathrm{E}\left[\pi_{\gamma k} | Y_k = y\right]$ is unknown, there are two natural ways to estimate it by using the sampling weights when they are available. The first way to estimate ρ_∞ consists of using the Hájek-type kernel estimator of $\mathrm{E}\left[\pi_{\gamma k} | Y_k = y\right]$:

$$\hat{\rho}_\infty = \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)\right)^{-1} N_\gamma \frac{\sum_{k=1}^{N_\gamma} \pi_{\gamma k} K\left(\frac{Y_k - y_0}{h_\gamma}\right) \left(\left(\mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)\right)/\pi_{\gamma k}\right)}{\sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) \left(\left(\mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)\right)/\pi_{\gamma k}\right)}.$$

The second way consists of using the inverse of the Horvitz-Thompson plug-in kernel estimator of $\mathrm{E}\left[\frac{1}{\pi_{\gamma k}} | Y_k = y\right]$:

$$\tilde{\rho}_\infty = \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)\right)^{-1} N_\gamma \frac{\sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) \left(\left(\mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)\right)/\pi_{\gamma k}\right)}{\sum_{k=1}^{N_\gamma} \frac{1}{\pi_{\gamma k}} K\left(\frac{Y_k - y_0}{h_\gamma}\right) \left(\left(\mathbb{1}_{\mathbb{N}\setminus\{0\}}(I_k)\right)/\pi_{\gamma k}\right)}.$$

Recall that:

$$p_\gamma(y_0) = \frac{1}{h_\gamma \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)\right)} \sum_{k=1}^{N_\gamma} K\left(\frac{Y_k - y_0}{h_\gamma}\right) I_{\gamma k},$$

so that

$$\begin{aligned} \frac{p_\gamma(y_0)}{\hat{\rho}_\infty(y_0)} &= \left(N_\gamma \left(\sum_{k\in U_\gamma} \frac{I_{\gamma k}}{\pi_{\gamma k}}\right)^{-1}\right) \hat{f}_\gamma(y_0) \\ &= \hat{f}_\gamma(y_0) \qquad \text{if} \quad \left(\sum_{k\in U_\gamma} \frac{I_{\gamma k}}{\pi_{\gamma k}}\right) = N_\gamma. \end{aligned} \tag{4.6}$$

Equation (4.6) means that if the selection is balanced on the population size, $\frac{\mathrm{p}_\gamma(y_0)}{\hat{\rho}_\infty(y_0)}$ is the Hájek type kernel density estimator of f.

4.4 Examples

We consider some examples from the series of examples studied in chapter 3. We give examples where conditions A4.3, A4.4, A4.5a hold and where they fail.

4.4.1 Non-informative selection

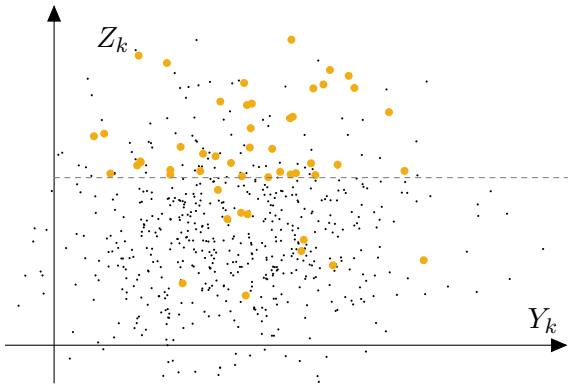
The following set of assumptions consist of a condition for the sample size to grow to ∞ (4.6a), classical assumptions (e.g. classical for the case of a census) for the kernel density estimator to be asymptotically

unbiased (4.6b and 4.6c), and for the variance of the kernel density estimator to tend to 0 when $N_\gamma h_\gamma$ tends to ∞ (4.6d):

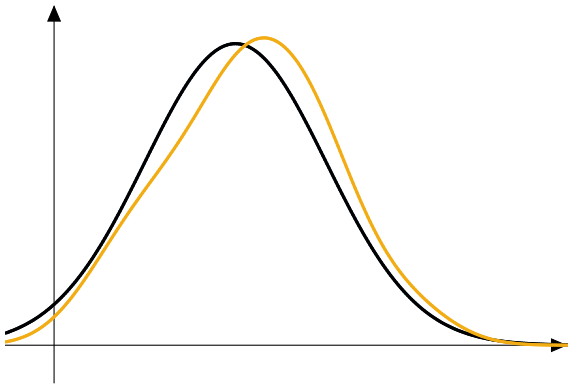
$$\bullet\mathbf{A4.6}\begin{cases}\operatorname{Var}[n_\gamma] = o_\gamma(N_\gamma^2) & (4.6a)\\ \sup_{y\in\mathbb{R}} f(y) < +\infty & (4.6b)\\ f \text{ is continuous in } y_0 & (4.6c)\\ \int K^2 \,\mathrm{d}\lambda < \infty & (4.6d)\end{cases}$$

In non-informative selection, A4.3 and A4.4 hold when A4.6 holds. In chapter 3, conditions for A3.0 and A3.1 have been given for a list of non-informative sample schemes of fixed or random size. Assumption A3.1 implies A4.6a.

Figure 4.1: Independent stratified sampling



(a) Y_k and Z_k are independent. The selection is non-informative, $\rho_\infty = 1$ and $\rho_\infty f = f$.



(b) Non-informative selection: the kernel density estimator (orange) is close to $\rho_\infty.f = f$ (black).

4.4.2 Cluster sampling

Here is presented an example of non convergence of the kernel density estimator to the limit sample pdf:

Suppose $Y_k \sim \mathcal{N}(\beta, 1)$, $\mathrm{P}^{\mathcal{Z}_\gamma|\mathcal{Y}_\gamma=y} = \mathcal{N}(\xi.y, \mathrm{Id}_{N_\gamma})$. For a vector z of $\mathbb{R}^N$, define $\zeta_\alpha(z)$ as the quantile of order α of the values of the vector z: $\zeta_\alpha(z) = \inf\left\{z_k \,|k \in U, N^{-1}\#\left\{\ell \in \{1,\dots,N\} \,|z_\ell \leqslant z_k\right\} \geqslant \alpha\right\}$. Assume the design measure function is the function characterized by:

$\forall z \in [0,1]^{N_\gamma}, i \in \mathbb{N}^{N_\gamma}$,

$$\begin{aligned}(D_\gamma(z)) : \{i\} \quad \mapsto \quad & \frac{1}{2}\left(\left(\prod_{k=1}^{N_\gamma}\left(\mathbb{1}_{[0,\zeta_{0.25}(z)]}(z_k)\right)^{i_k}\left(1-\mathbb{1}_{[0,\zeta_{0.25}(z)]}(z_k)\right)^{1-i_k}\right)\right.\\ & \left.+\left(\prod_{k=1}^{N_\gamma}\left(\mathbb{1}_{[\zeta_{0.75}(z),1]}(z_k)\right)^{i_k}\left(1-\mathbb{1}_{[\zeta_{0.75}(z),1]}(z_k)\right)^{1-i_k}\right)\right).\end{aligned}$$

(4.7) means that with probability $1/2$ the sample constituted of the elements that correspond to the 25% smallest values of $\mathcal{Z}_\gamma$ are selected, and with probability $1/2$ the sample constituted of the elements that correspond to the 25% largest values of $\mathcal{Z}_\gamma$ are selected.

Note that

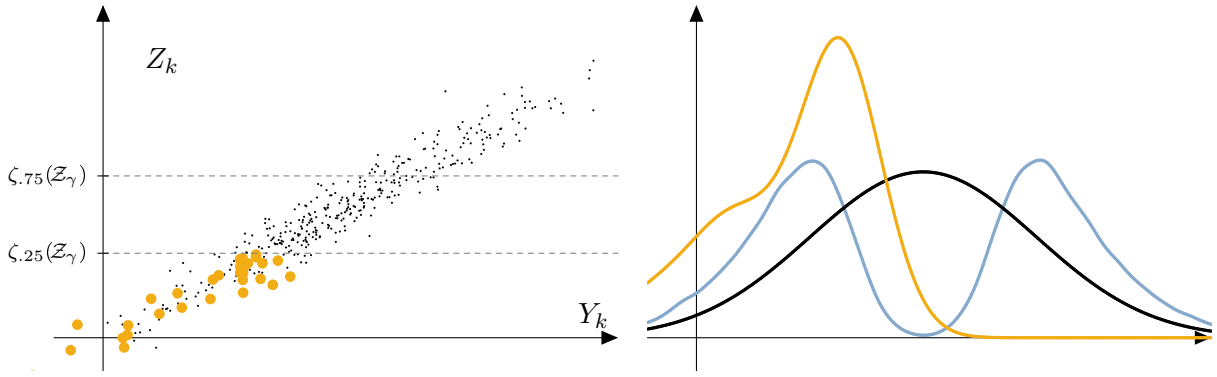
$$\begin{aligned}
&\mathrm{Cov}\left[I_{\gamma k}, I_{\gamma \ell} \mid Y_k = y_1, Y_\ell = y_2\right] \\
&= \left(\frac{1}{2}\,\mathrm{P}\left(Z_{\gamma_1} < \zeta_{0.25}(\mathcal{Z}_\gamma) \cap Z_{\gamma_2} < \zeta_{0.25}(\mathcal{Z}_\gamma) | Y_1 = y_1, Y_2 = y_2\right)\right. \\
&\qquad \left. + \frac{1}{2}\,\mathrm{P}\left(Z_{\gamma_1} > \zeta_{0.75}(\mathcal{Z}_\gamma) \cap Z_{\gamma_2} > \zeta_{0.75}(\mathcal{Z}_\gamma) | Y_1 = y_1, Y_2 = y_2\right)\right) \\
&\quad -\frac{1}{2}\left(\mathrm{P}\left(Z_{\gamma_1} < \zeta_{0.25}(\mathcal{Z}_\gamma) | Y_1 = y_1, Y_2 = y_2\right) + \mathrm{P}\left(Z_{\gamma_1}(\mathcal{Z}_\gamma) > \zeta_{0.75} | Y_1 = y_1, Y_2 = y_2\right)\right) \\
&\qquad \times \frac{1}{2}\left(\mathrm{P}\left(Z_{\gamma_1} < \zeta_{0.25}(\mathcal{Z}_\gamma) | Y_1 = y_1, Y_2 = y_2\right) + \mathrm{P}\left(Z_{\gamma_2}(\mathcal{Z}_\gamma) > \zeta_{0.75} | Y_1 = y_1, Y_2 = y_2\right)\right)
\end{aligned}$$

Denote $A(y) = \Phi\left(\sqrt{1+\xi^2}\Phi^{-1}(0.25) + \xi(\beta - y)\right)$, $B(y) = 1 - \Phi\left(\sqrt{1+\xi^2}\Phi^{-1}(0.75) + \xi(\beta - y)\right)$. Then,

$$\begin{aligned}
\lim_{\gamma\to\infty} c_\gamma(y_1, y_2) &= \left(\frac{1}{2}\left(A(y_1)A(y_2) + B(y_1)B(y_2)\right)\right) \\
&\qquad - \left(\frac{1}{2}\left(A(y_1) + B(y_1)\right)\right)\left(\frac{1}{2}\left(A(y_2) + B(y_2)\right)\right) \\
\lim_{\gamma\to\infty} d_\gamma(y_1, y_2) &= 0 \\
\lim_{\gamma\to\infty} (c_\gamma + d_\gamma)(y_1, y_2) &= \frac{1}{4}\left(A(y_1)(A(y_2) - B(y_2) + B(y_1)(B(y_2) - A(y_2))\right) \\
&\neq 0
\end{aligned}$$

so that A4.4f fails to hold.

Figure 4.2: Kernel density estimation in the case of cluster sampling



(a) At first stage, the first cluster has been selected.

(b) Plot of population pdf (black), kernel density estimator (orange), and sample pdf (blue).

4.4.3 With-replacement sampling with probability proportional to size

Let $\{n_\gamma^*\}$ be a non-random sequence of positive integers such that $\lim_{\gamma\to\infty} N_\gamma^{-1} n_\gamma^* = \tau \in]0, +\infty]$.

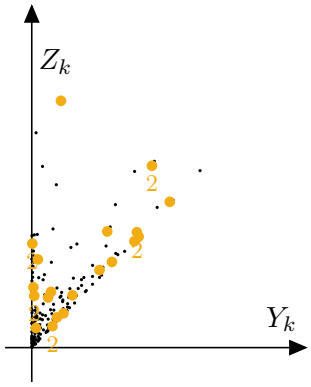
Assume $\mathcal{Z}_\gamma$ is a positive random variable. Assume $(Y_1, Z_{\gamma 1}), \ldots (Y_{N_\gamma}, Z_{\gamma N_\gamma})$ are iid couples of random variables, and that $\mathrm{P}^{(Y_1, Z_{\gamma 1})}$ does not depend on γ. Assume that $\forall z \in \mathbb{N}^{N_\gamma}$, $D_\gamma(z) = \mathrm{SWR}_{z, n_\gamma^*}$.

Under mild additional conditions (continuity of f, continuity of $y \mapsto \mathrm{E}\left[Z_{\gamma k}|Y_k = y\right]$, $\sup_{y\in\mathbb{R}} \mathrm{E}\left[Z_{\gamma k}|Y_k = y\right] f(y) < +\infty$, existence of finite moment of order 6 for $Z_{\gamma k}$) , A3.0, A3.1, A4.3, and A4.4 can be established using straightforward bounding and limiting arguments, like in section C.3.1, appendix C. Then,

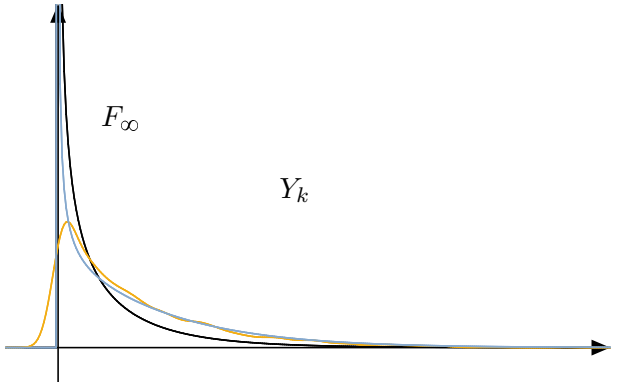
$$\rho_\infty(y) = \frac{\mathrm{E}\left[Z_{\gamma k}|Y_k = y\right]}{\mathrm{E}\left[Z_{\gamma k}\right]}$$
$$v_\infty(y_k) = \tau \frac{\mathrm{E}\left[Z_{\gamma k}|Y_k = y\right]}{\mathrm{E}\left[Z_{\gamma k}\right]}.$$

Assume for example that $Y_{\gamma 1} \sim \chi^2(1)$, $\mathrm{P}^{Z_{\gamma k}-y|Y_{\gamma k}=y} = \chi^2(1)$. In that case, it is possible to show that $\rho_\infty(y) = \frac{y+1}{2}$ and $\rho_\infty f(y) = \pi^{-1/2}\frac{1}{2}(y^{3/2} + y^{5/2})\, e^{-y}$.

Figure 4.3: Sampling with replacement and probability proportional to size



(a) Plot of $(Y_k, Z_{\gamma k})_{k\in U_\gamma}$, for $N_\gamma = 500$. The large circles correspond to units selected more than once, the numbers below indicate the number of times units were selected.



(b) Plot of the population pdf f (black), the sample kernel density estimator p_γ (orange) and the limit sample pdf $\rho_\infty f$ for $N_\gamma = 10000$. The limit sample pdf p is close to ρf

References

Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer Verlag.

Chapter 5

Sample likelihood estimation

In this chapter, we use the iid superpopulation model and the asymptotic framework described in chapter 3, and further conditions on the selection mechanism and on the sample pdf regularity. We assume a parametric model of the law of Y, parametrized by θ, and a parametric model of the distribution of Z given Y, parametrized by ξ. Assuming we have a consistent estimator of ξ, we plug this estimator into the approximate likelihood based on the sample distribution. We refer to this approximate likelihood as the sample likelihood. and we study the properties of the estimator of θ that maximizes the sample likelihood in θ. Adapting Gong and Samaniego (1981), we prove the existence, the consistency and the rate of convergence to a normal distribution of the maximum sample likelihood estimator of θ. When the design structure and observation mechanism are well known by the analyst, using maximum likelihood derived from the sample distribution is possible and can allow estimation with smaller variance than the pseudo-likelihood estimators derived from Horvitz-Thompson likelihood estimation. We illustrate this final result by the analysis of an example of stratified sampling, along with supporting simulations.

5.1 Notations and definitions

5.1.1 Assumptions

Let $(N_\gamma)_{\gamma\in\mathbb{N}}$ be a sequence of population sizes such that $\lim_{\gamma\to\infty} N_\gamma = +\infty$. All random variables are defined on a common measurable space $(\Omega, \mathscr{A})$, and we consider the statistical model $(\Omega, \mathscr{A}, \mathrm{P}_{\theta,\xi})_{(\theta,\xi)\in\Theta\times\Xi}$ where Θ and Ξ are open subsets of $\mathbb{R}$. Let $p,\ q \in \mathbb{N}\backslash\{0\}$. We assume that $(\mathscr{Y}, \mathscr{T}_Y) = (\mathbb{R}^q, \mathscr{B}_{\mathbb{R}^q})$. and $(\mathscr{Z}, \mathscr{T}_Z) = (\mathbb{R}^q, \mathscr{B}_{\mathbb{R}^q})$. We assume that $((Y_k, Z_k))_{k\in\mathbb{N}}$ is a sequence of independent and iid pairs of random variables, and $\forall\gamma \in \mathbb{N}$, $\mathcal{Y}_\gamma = (Y_1, \ldots, Y_{N_\gamma})$, $\mathcal{Z}_\gamma = (Z_1, \ldots, Z_{N_\gamma})$. We assume that $\mathrm{P}_{\theta,\xi}^{Y_k,Z_k}$ admits a density with respect to λ_{p+q}, the Lebesgue measure on $\mathbb{R}^{p+q}$. We assume that $\mathrm{P}_{\theta,\xi}^{Y_k}$ does not depend on ξ and denote by $f_\theta = d\,\mathrm{P}_{\theta,\xi}^{Y_k}/\mathrm{d}\lambda_p$ its density. Further we assume that $\mathrm{P}_{\theta,\xi}^{Z_k|Y_k}$ does not depend on θ.

Definition 5.1. *The sample pdf*

For $\gamma \in \mathbb{N}$, $\theta \in \Theta$, $\xi \in \Xi$, we define $m_{\gamma,\theta,\xi} : y \mapsto \mathrm{E}_{\theta,\xi}\left[I_{\gamma 1}|Y_1 = y\right]$. If $0 < \mathrm{E}_{\theta,\xi}\left[I_{\gamma k}\right] < +\infty$, then we can define the sample pdf as the function $\rho_{\gamma,\theta,\xi} f_\theta$, where

$$\begin{aligned}\rho_{\gamma,\theta,\xi} :\ \mathbb{R}^p &\to \mathbb{R}\\ y &\mapsto \rho_{\gamma,\theta,\xi}(y) = \frac{\mathrm{E}_{\theta,\xi}\left[I_{\gamma k}|Y_k = y\right]}{\mathrm{E}_{\theta,\xi}\left[I_{\gamma k}\right]}\\ &= \frac{m_{\gamma,\theta,\xi}}{\int m_{\gamma,\theta,\xi} f_\theta\,\mathrm{d}\lambda_p},\end{aligned}$$

and the sample distribution as the measure $\rho_{\gamma,\theta,\xi} f_\theta . \lambda_p$.

The following conditions on ρ_γ will allow us to define the limit sample pdf:

•**A5.0.**

$$\begin{cases} \forall \theta \in \Theta, \xi \in \Xi, \exists\, M_{\theta,\xi} : \mathbb{R}^p \to \mathbb{R}^+ \ \lambda\text{-measurable such that } \begin{cases} \forall \gamma \in \mathbb{N}, m_{\gamma,\theta,\xi} < M_{\theta,\xi}, \\ \int M_{\theta,\xi} f_\theta \ \mathrm{d}\lambda_p < \infty, \end{cases} & \text{(5.0a)} \\ \forall \theta \in \Theta, \xi \in \Xi, \exists\, m_{\infty,\theta,\xi} : \mathbb{R}^p \to \mathbb{R}^+ \ \lambda\text{-measurable such that } \begin{cases} m_{\gamma,\theta,\xi} \underset{\gamma\to\infty}{\to} m_{\infty,\theta,\xi} \text{ pointwise}, \\ \int m_{\infty,\theta,\xi} f_\theta \ \mathrm{d}\lambda_p > 0. \end{cases} & \text{(5.0b)} \end{cases}$$

Definition 5.2. *The limit sample pdf*
Under A5.0 we define the limit sample pdf: $\rho_{\infty,\theta,\xi} = \lim_{\gamma\to\infty} \rho_{\gamma,\theta,\xi}$.

5.1.2 Further definitions

Definition 5.3. *Given* $\gamma, (k,\ell) \in \{1,\dots,N_\gamma\}^2$, $k \neq \ell$, *we define the following functions:*

$$\begin{aligned} m_{\gamma,\theta,\xi} &: y \mapsto \mathrm{E}_{\theta,\xi}\left[I_{\gamma 1} | Y = y_1, Y_\ell = y_2\right], \\ v_{\gamma,\theta,\xi} &: y \mapsto \mathrm{Var}_{\theta,\xi}\left[I_{\gamma k} | Y_k = y\right], \\ m'_{\gamma,\theta,\xi} &: y_1, y_2 \mapsto \mathrm{E}_{\theta,\xi}\left[I_{\gamma \ell} | Y_k = y_1, Y_\ell = y_2\right], \\ c_{\gamma,\theta,\xi} &: y_1, y_2 \mapsto \mathrm{Cov}_{\theta,\xi}\left[I_{\gamma k}, I_{\gamma \ell} | Y_k = y_1, Y_\ell = y_2\right], \\ d_{\gamma,\theta,\xi} &: y_1, y_2 \mapsto m'_{\gamma,\theta,\xi}(y_1,y_2) m'_{\gamma,\theta,\xi}(y_2,y_1) - m_{\gamma,\theta,\xi}(y_1) m_{\gamma,\theta,\xi}(y_2). \end{aligned}$$

These definitions do not depend on the choice of k, ℓ *under the exchangeability conditions* (2.4) *and* (2.8).

5.1.3 Assumptions on the design measure: asymptotic independence of draws

•**A5.1.** L_2 **convergence conditions:** *Let* $\theta \in \Theta, \xi \in \Xi$ *and* $u : (\mathbb{R}^p, \mathscr{B}_{\mathbb{R}^p}) \to (\mathbb{R}^q, \mathscr{B}_{\mathbb{R}^q})$ *a measurable function.*

$$\begin{cases} A5.0 \text{ is satisfied and } \int \|u\| M_{\theta,\xi} f_\theta \lambda_p < +\infty, & \text{(5.1a)} \\ \int |u|(y_1) \, \|u\|(y_2) c_{\gamma,\theta,\xi}(y_1,y_2) f_\theta(y_1) f_\theta(y_2) . \, \mathrm{d}\lambda_p(y_1) \, \mathrm{d}\lambda_p(y_2) = o_\gamma(1), & \text{(5.1b)} \\ \int |u|(y_1) \, \|u\|(y_2) d_{\gamma,\theta,\xi} f_\theta(y_1) f_\theta(y_2) . \, \mathrm{d}\lambda_p(y_1) \, \mathrm{d}\lambda_p(y_2) = o_\gamma(1), & \text{(5.1c)} \\ \int \|u\|^2 (v_{\gamma,\theta,\xi} + \rho^2_{\gamma,\theta,\xi}) f_\theta \ \mathrm{d}\lambda_p = o_\gamma(N_\gamma), & \text{(5.1d)} \\ \mathrm{P}_{\theta,\xi}\left(\mathcal{I}_\gamma = 0\right) = o_\gamma(1), & \text{(5.1e)} \end{cases}$$

where $\|.\|$ *is the Euclidian norm on* $\mathbb{R}^q$, *and* $\|u\| : x \mapsto \|u(x)\|$.

5.2 Maximum sample likelihood estimation

The following series of results follows closely the results of Gong and Samaniego (1981), adapted to the context of informative selection which we have described above.

5.2.1 Approximation of log-likelihood based on the sample distribution

Definition 5.4. *Define* $\Delta : \mathbb{R}^p \times \Theta \times \Xi \to \mathbb{R}, (y, \theta, \xi) \mapsto \Delta(y, \theta, \xi) = \ln\left(\rho_{\infty,\theta,\xi}(y) f_\theta(y)\right)$.

For $\gamma \in \mathbb{N}$, *we define the mean log sample likelihood as*

$$\begin{aligned}
&\overline{\mathscr{L}_\gamma}\left(\theta, \xi, \left(Y_{R_\gamma(k)}\right)_{k\in\{1,\dots,n_\gamma\}}\right) \\
&= \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)\right)^{-1} \sum_{k=1}^{n_\gamma} \Delta\left(Y_{R_\gamma(k)}, \theta, \xi\right) \\
&= \left(n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)\right)^{-1} \sum_{k=1}^{N_\gamma} I_{\gamma k} \Delta\left(Y_k, \theta, \xi\right),
\end{aligned}$$

and the maximum sample likelihood estimator of θ *adapted to* ξ *is*

$$\hat{\theta}_\gamma(\xi) = \arg\max_{\theta\in\Theta} \left\{ \overline{\mathscr{L}_\gamma}\left(\theta, \xi, \left(Y_{R_\gamma(k)}\right)_{k\in\{1,\dots,n_\gamma\}}\right)\right\}.$$

5.2.2 Maximum likelihood estimation based on the sample distribution

Lemma 5.1. *We assume A5.0. Let* $\theta_0 \in \Theta$, $\xi_0 \in \Xi$. *Let* $\hat{\xi}_\gamma$ *be a random variable such that* $\hat{\xi}_\gamma - \xi_0 = o_{\mathrm{P}_{\theta_0,\xi_0}}(1)$. *Let* $\Lambda : \mathbb{R}^p \times \Xi \to \mathbb{R}, (y, \xi) \mapsto \Lambda(y, \xi)$, *and let* B *be a neighborhood of* ξ_0 *such that* $\lambda_p - a.s.(y)$, $\Lambda(y, .)$ *is a differentiable function of* ξ *for* ξ *in* B. *We assume that* $\exists R : (\mathbb{R}^p, \mathscr{B}_{\mathbb{R}^p}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ *a measurable function such that* $\forall \xi \in B$, $\left|\frac{\partial}{\partial\xi}\Lambda(y, \xi)\right| \leqslant R(y)$. *We assume that A5.1 is satisfied at* θ_0, ξ_0 *for* $u = \mathbb{1}_{\mathbb{R}^p}$, $\Lambda(., \xi_0)$, *and* R. *Then*

$$\frac{\sum_{k=1}^{N_\gamma} I_{\gamma k} \Lambda(Y_k, \hat{\xi}_\gamma)}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} \xrightarrow[\gamma\to\infty]{\mathrm{P}_{\theta_0,\xi_0}} \int \Lambda(y, \xi_0)\, \rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y)\, \mathrm{d}\lambda(y).$$

Proof. See appendix D.1. □

•**A5.2.** *We will say that A5.2 is satisfied for* $\theta_0 \in \Theta$, $\xi_0 \in \Xi$, A *a neighborhood of* θ_0 *and* B *a neighborhood of* ξ_0 *if*

$$A5.1 \text{ is satisfied for } u = \mathbb{1}_{\mathbb{R}^p}, \tag{5.2a}$$

$$\lambda_p - a.s.(y), \forall \theta \in A, \xi \in B, \frac{\partial}{\partial\theta}\Delta, \frac{\partial^2}{\partial\theta^2}\Delta, \frac{\partial^3}{\partial\theta^3}\Delta, \frac{\partial}{\partial\xi}\Delta, \frac{\partial^2}{\partial\theta\partial\xi}\Delta, \frac{\partial^3}{\partial\theta^2\partial\xi}\Delta, \frac{\partial^3}{\partial\theta\partial\xi^2}\Delta \text{ are defined,} \tag{5.2b}$$

Interchange of differentiation and integration of $\rho_{\infty,\theta,\xi} f_\theta$ *is valid for first and second derivatives with respect to* θ *and for the mixed partial derivative with respect to* θ *and* ξ, (5.2c)

$$0 < \mathscr{I}_{11} = \int_{\mathbb{R}^p} \left(\frac{\partial\Delta}{\partial\theta}(y, \theta_0, \xi_0)\right)^2 d\left(\rho_{\infty,\theta_0,\xi_0} f_{\theta_0}.\lambda_p\right)(y) < +\infty,$$

$$\int_{\mathbb{R}^p} \left|\left(\frac{\partial\Delta}{\partial\xi}\frac{\partial\Delta}{\partial\theta}\right)(y, \theta_0, \xi_0)\right| d\left(\rho_{\infty,\theta_0,\xi_0} f_{\theta_0}.\lambda_p\right)(y) < +\infty,$$

$$\text{and then define } \mathscr{I}_{12} = \int_{\mathbb{R}^p} \left(\frac{\partial\Delta}{\partial\xi}\frac{\partial\Delta}{\partial\theta}\right)(y, \theta_0, \xi_0)\, d\left(\rho_{\infty,\theta_0,\xi_0} f_{\theta_0}.\lambda_p\right)(y), \tag{5.2d}$$

$$\exists K : \mathbb{R}^p \times A \to \mathbb{R}^+ \text{ such that } K(., \theta) \text{ satisfies A5.1 and } \forall \theta, \xi \in A \times B, y \in \mathbb{R}^p$$

$$\left|\frac{\partial}{\partial\xi}\ln\left(\frac{\rho_{\infty,\theta,\xi}(y)f_\theta(y)}{\rho_{\infty\theta_0\xi}(y)f_{\theta_0}(y)}\right)\right| < K(y,\theta), \tag{5.2e}$$

$$\exists L:\mathbb{R}^p\times A\to\mathbb{R}^+ \text{ a measurable function such that}$$

$$\forall\theta\in A,\xi\in B,\lambda_p-a.s(y),\left|\frac{\partial^3\Delta}{\partial\theta^3}(y,\theta,\xi)\right|\leqslant L(y),\left|\frac{\partial^3\Delta}{\partial\theta^2\partial\xi}(y,\theta_0,\xi)\right|\leqslant L(y),\left|\frac{\partial^3\Delta}{\partial\theta\partial\xi^2}(y,\theta_0,\xi)\right|\leqslant L(y),$$

$$\text{and }\int L(y)\rho_{\infty,\theta_0,\xi_0}(y)f_{\theta_0}(y)\,\mathrm{d}\lambda_p(y)<\infty, \tag{5.2f}$$

$$\forall(\theta,\xi)\in\Theta\times\Xi \text{ such that } (\theta,\xi)\neq(\theta_0,\xi_0),\int \mathbb{1}_{\{0\}}\left(\Delta(y,\theta,\xi)-\Delta(y,\theta_0,\xi_0)\right)\rho_{\infty,\theta_0,\xi_0}(y)f_{\theta_0}(y)\lambda_p(y)<1, \tag{5.2g}$$

$$\left(\frac{\partial\Delta}{\partial\theta}(.,\theta_0,\xi_0)\right)^2 \text{ and } \left(\frac{\partial\Delta}{\partial\theta}(.,\theta_0,\xi_0)\frac{\partial\Delta}{\partial\xi}(.,\theta_0,\xi_0)\right) \text{ satisfy A5.1 for } \theta_0,\xi_0. \tag{5.2h}$$

,

Theorem 5.1. *Existence of a consistent root of the ML equation*

We assume A0. Let $\theta_0\in\Theta$, $\xi_0\in\Xi$. Let A (resp. B) be a neighborhood of θ_0 (resp. ξ_0). We assume A5.2 is satisfied for θ_0,ξ_0,A,B. Let $\hat{\xi}_\gamma$ be a random variable such that $\hat{\xi}_\gamma-\xi_0=o_{\mathrm{P}_{\theta_0,\xi_0}}(1)$. For $\gamma\in\mathbb{N},\varepsilon\in\mathbb{R}^+$, let $C_\gamma(\varepsilon)$ be the event that there exists a root $\hat{\theta}_\gamma$ to the equation: $\sum_{k=1}^{N_\gamma}\left(\partial\Delta/\partial\theta\right)\left(Y_k,\theta,\hat{\xi}_\gamma\right)I_{\gamma k}=0$ such that $\left|\hat{\theta}_\gamma-\theta_0\right|<\varepsilon$. Then, $\lim_{\gamma\to\infty}\mathrm{P}_{\theta_0,\xi_0}\left(C_\gamma(\varepsilon)\right)=1$.

Proof. See appendix D.2. □

Theorem 5.2. *We assume A5.0. Let $\theta_0\in\Theta$, $\xi_0\in\Xi$. Let A (resp. B) be a neighborhood of θ_0 (resp. ξ_0). We assume that A5.2 is satisfied for θ_0,ξ_0,A,B. We assume θ_0,ξ_0 is the true parameter. Let $\hat{\xi}_\gamma$ be random variable such that $\hat{\xi}_\gamma-\xi_0=O_{\mathrm{P}_{\theta_0,\xi_0}}\left(\frac{1}{\sqrt{N_\gamma}}\right)$. Suppose $\sqrt{n_\gamma}\begin{bmatrix}\left(\frac{\partial}{\partial\theta}\overline{\mathscr{L}}\right)\left(\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\dots,n_\gamma\}},\theta_0,\xi_0\right)\\ \hat{\xi}_\gamma-\xi_0\end{bmatrix}\xrightarrow[\gamma\to\infty]{\mathscr{L}}\mathscr{N}\left(0,\Sigma=\begin{bmatrix}\Sigma_{11}&\Sigma_{12}\\ &\Sigma_{22}\end{bmatrix}\right)$. Then the maximum sample likelihood estimator adapted to $\hat{\xi}_\gamma$ is asymptotically normal, that is*

$$\sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)/\sigma\xrightarrow[\gamma\to\infty]{\mathscr{L}}\mathscr{N}(0,1),$$

where

$$\sigma^2 = \frac{\Sigma_{11}}{\mathscr{I}_{11}^2}+\frac{\mathscr{I}_{12}}{\mathscr{I}_{11}^2}\left(\Sigma_{22}\mathscr{I}_{12}-2\Sigma_{12}\right).$$

Proof. See appendix D.3. □

5.3 Example: stratified sampling

5.3.1 Asymptotic framework

Let $H\in\mathbb{N}\backslash\{0\}$ be a fixed and non-random number of strata, and let $(N_{\gamma h})_{\gamma\in\mathbb{N},h\in\{1,\dots,H\}}$ be an array of strictly positive integers such that $\forall\gamma\in\mathbb{N}$, $N_\gamma=\sum_{h=1}^H N_{\gamma h}$. For $\gamma\in\mathbb{N},h\in\{1,\dots,H\}$, $N_{\gamma h}$ denotes the size of the hth stratum of the γth population. For $\gamma\in\mathbb{N}$, $h\in\{1,\dots,H\}$, let $n_{\gamma h}\in\{1,\dots,N_{\gamma h}\}$ denote the number of elements selected from the hth stratum of the γth population. We define ν_γ the

permutation such that $Z_{\nu_\gamma(1)} < \cdots < Z_{\nu_\gamma(N_\gamma)}$. The permutation ν_γ is a random variable which is a function of $\mathcal{Z}_\gamma$. The hth stratum of the γth population is the set: $U_{\gamma h} = \nu_\gamma\left(\{T_{\gamma h-1}+1,\ldots,T_{\gamma h}\}\right)$, with $T_{\gamma 0}=0$, $T_{\gamma h} = \sum_{1\leqslant h'\leqslant,h} N_{\gamma h'}$. For $h \in \{0,\ldots,H\}$, define $t_{\gamma h} = \frac{T_{\gamma h}}{N_\gamma}$. The random design measure is stratified simple random sampling without replacement:

$$\Pi_\gamma(\{i\}) = \begin{vmatrix} \prod_{h=1}^{H} \binom{N_{\gamma h}}{n_{\gamma h}}^{-1} & \text{if } \forall h \in \{1,\ldots,H\}, \sum_{k\in U_{\gamma h}} i_k = n_{\gamma h}, \text{ and } i \in \{0,1\}^{N_\gamma} \\ 0 & \text{otherwise.} \end{vmatrix}$$

We assume that $\forall h \in \{0,\ldots,H\}$, $t_{\infty,h} = \lim_{\gamma\to\infty} t_{\gamma h}$ is defined. We also assume that $\forall h \in \{1,\ldots,H\}$, $\tau_h = \lim_{\gamma\to\infty} N_{\gamma h}^{-1} n_{\gamma h}$ is defined. We assume that $\tau = \sum_{h=1}^{H} \tau_h \left(t_{\infty,h} - t_{\infty,h-1}\right) = \lim_{\gamma\to\infty} N_\gamma^{-1} n_\gamma^* > 0$, with $n_\gamma^* = \sum_{k=1}^{H} n_{\gamma h}$. We assume that $Y_k \sim \mathcal{N}(\theta,1)$, $P^{Z_k|Y_k} = \mathcal{N}(\xi.Y_k, \sigma^2)$, where σ is known, i.e $Z_k = \xi.Y_k + \sigma.\eta_k$ and $\eta_k \sim \mathcal{N}(0,1)$.

5.3.2 Maximum sample likelihood estimator

Result 5.1. *Under the asymptotic framework described in section 5.3.1, the assumption A5.0 holds, and*

$$\begin{aligned}
\rho_{\infty,\theta,\xi}(y) &= \left(\sum_{h=1}^{H} \tau_h \left(t_{\infty,h} - t_{\infty,h-1}\right)\right)^{-1} \\
&\quad \left(\sum_{h=1}^{H} \tau_h \left(\Phi\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y)\right)\right.\right. \\
&\quad \left.\left. -\Phi\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h-1}\right) + \frac{\xi}{\sigma}(\theta - y)\right)\right)\right) \\
&= \left(\tau_H + \sum_{h=1}^{H-1} t_{\infty,h}\left(\tau_h - \tau_{h+1}\right)\right)^{-1} \\
&\quad \left(\tau_H + \sum_{h=1}^{H-1} \left(\tau_h - \tau_{h+1}\right) \Phi\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y)\right)\right),
\end{aligned}$$

where Φ is the cdf of $\mathcal{N}(0,1)$.

Proof. See appendix D.4.1. □

The sample likelihood is then defined and if we consider

$$\hat{\xi}_\gamma = \frac{\sum_{k=1}^{n_\gamma} Z_{R_\gamma(k)} Y_{R_\gamma(k)} / \pi_{\gamma k}}{\sum_{k=1}^{n_\gamma} Y_{R_\gamma(k)}^2 / \pi_{\gamma k}},$$

we can define $\hat{\theta}_\gamma$ to be the maximum sample likelihood estimator of θ adapted to $\hat{\xi}_\gamma$, i.e.

$$\hat{\theta}_\gamma = \arg\min_{\theta\in\Theta} \left\{ \overline{\mathscr{L}}\left(\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}}, \theta, \hat{\xi}_\gamma\right)\right\}.$$

Result 5.2. *The statistic $\hat{\xi}_\gamma$ is a consistent estimator of ξ.*

Proof. See Appendix D.4.4. □

5.3.3 Existence of a consistent root to the MLE equation and limiting variance

Result 5.3. *Under the asymptotic framework described in section 5.3.1, the conditions of Theorem 5.1 are satisfied.*

Proof. See appendix D.4.2. □

Result 5.4. *Under the asymptotic framework described in section 5.3.1, the conditions of Theorem 5.2 are satisfied, we have*

$$\begin{aligned}
\Sigma_{11} &= \mathscr{I}_{11} \\
&= \tau^{-1} \sum_{h=1}^{H} \left(t_{\infty h} - t_{\infty h-1}\right) \tau_{\infty h} \operatorname{Var}_{\theta_0,\xi_0} \left[\left(\frac{\partial}{\partial \theta} \Delta \right) \left(Y_1, \theta_0, \xi_0\right) | Z_1 \in \left]\zeta(t_{\infty h-1}), \zeta(t_{\infty h})\right] \right] \\
\Sigma_{22} &= \tau^{-1} \sum_{h=1}^{H} \left(t_{\infty h} - t_{\infty h-1}\right) \tau_{\infty h}^{-1} \\
&\qquad \left[(\theta_0^2+1)^{-1} \quad \xi_0(\theta_0^2+1)^{-1}\right] \operatorname{Var}_{\theta_0,\xi_0} \left[\left[\begin{matrix} Y_1 Z_1 \\ Y_1^2 \end{matrix} \right] \middle| Z_1 \in \left]\zeta(t_{\infty h-1}), \zeta(t_{\infty h})\right] \right] \left[\begin{matrix} (\theta_0^2+1)^{-1} \\ \xi_0(\theta_0^2+1)^{-1} \end{matrix} \right]
\end{aligned}$$

$$\begin{aligned}
\Sigma &= \tau^{-1} \sum_{h=1}^{H} \left(t_{\infty h} - t_{\infty h-1}\right) \tau_{\infty h} \\
&\qquad \left[\begin{matrix} 1 & 0 & 0 \\ 0 & \frac{1}{\theta_0^2+1} & \frac{\xi_0}{\theta_0^2+1} \end{matrix} \right] \operatorname{Var}_{\theta_0,\xi_0} \left[\left[\begin{matrix} \left(\frac{\partial}{\partial \theta} \Delta\right)\left(Y_1, \theta_0, \xi_0\right) \\ Y_1 Z_1 / \tau_{\infty h} \\ Y_1^2 / \tau_{\infty h} \end{matrix} \right] \middle| Z_1 \in \left]\zeta(t_{\infty h-1}), \zeta(t_{\infty h})\right] \right] \left[\begin{matrix} 1 & 0 \\ 0 & \frac{1}{\theta_0^2+1} \\ 0 & \frac{\xi_0}{\theta_0^2+1} \end{matrix} \right],
\end{aligned}$$

$$\begin{aligned}
&\left(\frac{\partial}{\partial \theta} \Delta \right) \left(Y_1, \theta_0, \xi_0\right) \\
&= (y-\theta) + \frac{\sum_{h=1}^{H-1} \left(\tau_h - \tau_{h+1}\right) \frac{\xi}{\sigma} f_0 \left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y) \right)}{\tau_H + \sum_{h=1}^{H-1} \left(\tau_h - \tau_{h+1}\right) \Phi \left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y) \right)},
\end{aligned}$$

and

$$\begin{aligned}
&\left(\frac{\partial}{\partial \xi} \Delta \right) \left(Y_1, \theta_0, \xi_0\right) \\
&= \frac{\sum_{h=1}^{H-1} \left(\tau_h - \tau_{h+1}\right) \left(\frac{\xi/\sigma^2}{\sqrt{\frac{\xi^2}{\sigma^2}+1}} \Phi^{-1}\left(t_{\infty,h}\right) + \frac{(\theta - y)}{\sigma} \right) f_0 \left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y) \right)}{\tau_H + \sum_{h=1}^{H-1} \left(\tau_h - \tau_{h+1}\right) \Phi \left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2 + 1}\, \Phi^{-1}\left(t_{\infty,h}\right) + \frac{\xi}{\sigma}(\theta - y) \right)}.
\end{aligned}$$

Proof. See appendix D.4.3. □

We evaluate the value of the matrix Σ by Monte Carlo in section 5.3.5.

5.3.4 Comparison to other estimators

Let $\tilde{\theta}_\gamma = \left(\sum_{k=1}^{n_\gamma} Y_{R_\gamma(k)}/\pi_{\gamma k}\right) / \left(\sum_{k=1}^{n_\gamma} 1/\pi_{\gamma k}\right)$. As we are in a case of fixed size without replacement sampling, $\sum_{k=1}^{n_\gamma} 1/\pi_{\gamma k}$ is not random and $\sum_{k=1}^{n_\gamma} 1/\pi_{\gamma k} = N_\gamma$.

Result 5.5. *Under the asymptotic framework described in section 5.3.1,*

$$\sqrt{n_\gamma}\left(\tilde{\theta}_\gamma - \theta\right) \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathcal{N}\left(0, \tau \sum_{h=1}^{H} \left(t_{\infty h} - t_{\infty h-1}\right) \tau_{\infty h}^{-1} \operatorname{Var}\left[Y_1 \middle| Z_1 \in \left]\zeta(t_{\infty h-1}), \zeta(t_{\infty h})\right]\right]\right),$$

where for $h \in \{1, \ldots, H\}$,

$$\begin{aligned}
&\operatorname{Var}\left[Y_1 \middle| Z_1 \in \left]\zeta(t_{\infty h-1}), \zeta(t_{\infty h})\right]\right] \\
&= \left(t_{\infty h} - t_{\infty,h-1}\right)^{-1} \int y^2 \left(\mathbb{1}_{]B_{h-1},B_h]}\left(\xi y + \sigma\eta\right)\right) \frac{1}{2\pi\sigma} \exp\left(\frac{-(y-\theta)^2}{2}\right) \exp\left(\frac{-\eta^2}{2\sigma^2}\right) \mathrm{d}y \, \mathrm{d}\eta \\
&\quad - \left(\left(t_{\infty h} - t_{\infty,h-1}\right)^{-1} \int y \left(\mathbb{1}_{]B_{h-1},B_h]}(\xi y + \sigma\eta)\right) \frac{1}{2\pi\sigma} \exp\left(\frac{-(y-\theta)^2}{2}\right) \exp\left(\frac{-\eta^2}{2\sigma^2}\right) \mathrm{d}y \, \mathrm{d}\eta\right)^2,
\end{aligned}$$

with $B_h = \xi\theta + \sqrt{\xi^2 + \sigma^2}\Phi^{-1}\left(t_{\infty h}\right)$.

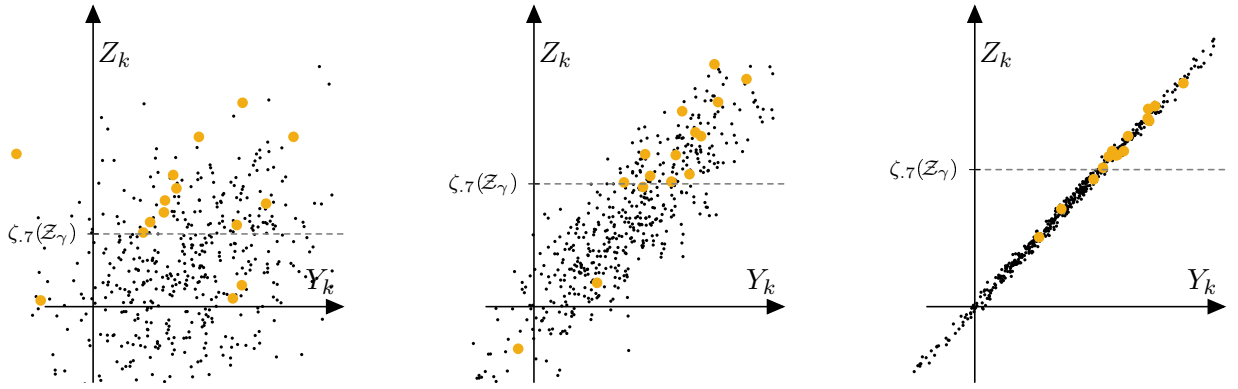
Proof. See appendix D.4.6. □

We evaluate these expressions by Monte Carlo in section 5.3.5.

5.3.5 Simulations

The following figure illustrates the realization of the population and the realization of the sample for different values of σ.

Figure 5.1: Different degrees of informative selection



(a) $\sigma = 10$, the selection is weakly informative

(b) $\sigma = 1$

(c) $\sigma = 0.1$, the selection is highly informative

For the simulations, we use the model described in section 5.3.1.

We generate $\alpha = 1000$ populations of $N_\gamma = 5000$ elements each, i.e. 1000 vectors of 5000 iid realizations of the distribution $\mathcal{N}(\theta, 1)$, with $\theta = 1.5$. We denote $(\mathcal{Y}_\gamma^{(\ell)})_{l\in\{1,\ldots,\alpha\}}$ the sequence of vectors. Then, for $\sigma \in \{0.1, 1, 10\}$, and for each vector $l \in \{1, \ldots, \alpha\}$, we create $\mathcal{Z}_\gamma^{(\ell)} = \xi.\mathcal{Y}_\gamma^{(\ell)} + \sigma.\eta^{(\ell)}$, where

$(\eta^{(\ell)})_{l\in\{1,\dots,\alpha\}}$ is a sequence of iid vectors, with $\eta^{(\ell)} \sim \mathcal{N}(0, \mathrm{Id}_{5000})$. Then for each $l \in \{1,\dots,\alpha\}$, we draw $\mathcal{I}_\gamma^{(\ell)}$ according to $\Pi^{(\ell)} = D(\mathcal{Z}_\gamma^{(\ell)})$, where D_γ is the design measure function that corresponds to the sample selection described in 5.3.1, with $H = 2$, $N_{\gamma 1} = 3500$, $N_{\gamma 2} = 1500$, $n_{\gamma 1} = 50$, $n_{\gamma 2} = 200$. We denote $\pi_{\gamma k}^{(\ell)}$ the vector of inclusion probabilies that correspond to $\Pi^{(\ell)}$. We compute $\rho_{\infty,\theta,\xi}$ by considering that $\lim_{\gamma\to\infty} N_{\gamma 1}^{-1} n_{\gamma 1} = 1/70$, $\lim_{\gamma\to\infty} N_{\gamma 2}^{-1} n_{\gamma 2} = 2/15$, $\lim_{\gamma\to\infty} N_\gamma^{-1} N_{\gamma 1} = 7/10$ and $\lim_{\gamma\to\infty} N_\gamma^{-1} N_{\gamma 2} = 3/10$. For all $l \in \{1,\dots,\alpha\}$, we compute

- $\hat{\xi}_\gamma^{(\ell)} = \left(\sum_{k=1}^{N_\gamma} \left(Y_k^{(\ell)}\right)^2 I_{\gamma k}^{(\ell)}/\pi_{\gamma k}^{(\ell)}\right)^{-1} \sum_{k=1}^{N_\gamma} Z_k^{(\ell)} I_{\gamma k}^{(\ell)} Y_k^{(\ell)}/\pi_{\gamma k}^{(\ell)}$,
- $\hat{\theta}_\gamma^{(\ell)} = \arg\max_{\theta\in\Theta} \left\{ \overline{\mathscr{L}_\gamma}\left(\theta, \hat{\xi}_\gamma^{(\ell)}, \left(Y^{(\ell)}_{R_\gamma^{(\ell)}(k)}\right)_{k\in\{1,\dots,n_\gamma\}}\right)\right\}$,
- $\tilde{\theta}_\gamma^{(\ell)} = \left(\sum_{k=1}^{N_\gamma} Y_k^{(\ell)} I_{\gamma k}^{(\ell)}/\pi_{\gamma k}^{(\ell)}\right) / \left(\sum_{\ell=1}^{N_\gamma} I_{\gamma k}^{(\ell)}/\pi_{\gamma k}^{(\ell)}\right)$,
- $\bar{\theta}_\gamma^{(\ell)} = (n_\gamma)^{-1} \sum_{k=1}^{N_\gamma} Y_k^{(\ell)} I_{\gamma k}^{(\ell)}$.

Then we compute

$$\mathrm{Mean}[\hat{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} \hat{\theta}_\gamma^{(\ell)}, \qquad \mathrm{Mean}[\bar{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} \bar{\theta}_\gamma^{(\ell)}, \qquad \mathrm{Mean}[\tilde{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} \tilde{\theta}_\gamma^{(\ell)},$$

$$\mathrm{MSE}[\hat{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} (\hat{\theta}_\gamma^{(\ell)} - \theta)^2, \quad \mathrm{MSE}[\bar{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} (\bar{\theta}_\gamma^{(\ell)} - \theta)^2, \quad \mathrm{MSE}[\tilde{\theta}] = \alpha^{-1}\sum_{\ell=1}^{\alpha} (\tilde{\theta}_\gamma^{(\ell)} - \theta)^2.$$

We repeat the operation three times, for three different values of σ: $\sigma \in \{0.1, 1, 10\}$. Independently, by Monte Carlo simulations, with 30000 realisations, we compute for the different values of σ the limiting variances of $\sqrt{n_\gamma}\hat{\theta}_\gamma$, $\sqrt{n_\gamma}\tilde{\theta}_\gamma$ and $\sqrt{n_\gamma}\bar{\theta}_\gamma$. The following table summarizes the results of the simulations and allows comparison of the theoretical variances to the observed mean square deviations: the estimator that maximizes the approximate likelihood has the smallest mean square error.

Table 5.1: Mean square error of different estimators

$N_\gamma = 5000$, $H = 2$, $N_{\gamma 1} = 3500$, $N_{\gamma 2} = 1500$, $n_{\gamma 1} = 50$, $n_{\gamma 1} = 200$, $\theta = 1.5$, $\xi = 2$, $\sigma \in \{0.1, 1, 10\}$, $\alpha = 1000$

θ	ξ	σ		Mean[.]	MSE[.]	$\sqrt{\frac{\mathrm{MSE}}{\mathrm{MSE}(\hat{\theta})}}$	$\frac{1}{n_\gamma}\lim_{\gamma\to\infty} n_\gamma \mathrm{Var}[.]$
1.5	2	0.1	$\hat{\theta}$	1.502	$7.643\ 10^{-4}$	1	$2.962\ 10^{-2}$
			$\tilde{\theta}$	1.5	$4.811\ 10^{-3}$	2.509	$1.023\ 10^{-2}$
			$\bar{\theta}$	2.329	$6.887\ 10^{-1}$	30.02	$3.979\ 10^{-3}$
1.5	2	1	$\hat{\theta}$	1.5	$1.975\ 10^{-3}$	1	$2.975\ 10^{-2}$
			$\tilde{\theta}$	1.501	$5.583\ 10^{-3}$	1.681	$1.024\ 10^{-2}$
			$\bar{\theta}$	2.241	$5.509\ 10^{-1}$	16.7	$3.971\ 10^{-3}$
1.5	2	10	$\hat{\theta}$	1.497	$5.501\ 10^{-3}$	1	$2.943\ 10^{-2}$
			$\tilde{\theta}$	1.5	$1.03\ 10^{-2}$	1.368	$1.03\ 10^{-2}$
			$\bar{\theta}$	1.662	$2.999\ 10^{-2}$	2.335	$4.027\ 10^{-3}$

References

Gong, G. and Samaniego, F. J. (1981). Pseudomaximum likelihood estimation: theory and applications. *The Annals of Statistics*, 9(4):861–869.

Chapter 6

Optimal inclusion probabilities for balanced sampling

When auxiliary information is available at the design stage, samples may be selected by means of balanced sampling. The variance of the Horvitz-Thompson estimator is then reduced, since it is approximately given by that of the residuals of the variable of interest on the balancing variables. In this chapter, two methods for computing optimal inclusion probabilities for balanced sampling on given auxiliary variables are studied. We show that the method formerly suggested by Tillé and Favre (2005) enables the computation of inclusion probabilities that lead to a decrease in variance under some conditions on the set of balancing variables. A disadvantage is that the target optimal inclusion probabilities depend on the variable of interest. If the needed quantities are unknown at the design stage, we propose to use estimates instead (e.g., arising from a previous wave of the survey). A limited simulation study suggests that our method performs well as compared to that suggested by Tillé and Favre (2005).

6.1 Introduction

A sampling design is said to be balanced if it leads to the selection of samples such that the Horvitz-Thompson estimators of the totals for auxiliary variables exactly match the known population totals. Many partial solutions were proposed for balanced sampling, before Deville and Tillé (2004) introduced the cube method. This sampling algorithm enables the selection of balanced samples with any number of balancing variables, and any prescribed set of inclusion probabilities.

Balanced sampling designs do not substitute for other classical and efficient sampling techniques, such as unequal probability sampling for selecting primary sampling units (PSUs) in household surveys, or stratification in business surveys. They may be thought of as a way to refine these techniques and obtain a variance reduction, if auxiliary information is available at the design stage. In this chapter, we propose to compute optimal inclusion probabilities for balanced sampling designs by means of a fixed-point algorithm, previously suggested by Tillé and Favre (2005).

Under some conditions on the set of balancing variables, we give a proof of convergence of this algorithm to a set of inclusion probabilities which correspond to a local minimum of the approximated variance. We thus propose to iterate the algorithm until having almost reached the limit. Whereas several iterations of the fixed-point algorithm are usually needed to get the target inclusion probabilities, we note that the set of inclusion probabilities obtained after one iteration is close to the final one. Consequently, considering only one iteration appears as a good trade-off between accuracy and simplicity. A disadvantage of the studied method is that some knowledge on the variable of interest is required, since quantities depending on the

variable of interest are needed for the fixed-point algorithm. If these quantities are unknown at the design stage, we propose to use estimated arising from another survey instead. Also, we note that the computed inclusion probabilities aim at minimizing, in some sense, an approximation of the variance. Consequently, it seems of interest to evaluate the performance of the proposed method with respect to the exact variance, through a limited simulation study. Our simulation results suggest that the proposed method performs well, as compared to the approximation originally proposed by Tillé and Favre (2005).

This chapter is organized as follows. The notation is defined in Section 6.2. A first algorithm for computing optimal inclusion probabilities is described in Section 6.3, and its properties are discussed. A limited simulation study is proposed in Section 6.4. A second algorithm is proposed for the case where the auxiliary variables are functions of the inclusion probabilities in Section 6.5. Our main conclusions are drawn in Section 6.6.

6.2 Notation and balanced sampling

For the reader's convenience, we recall the general notations that will be useful in this chapter. Let U denote a finite labeled population of size N. A sample without replacement i from the population U is a vector in $\{0,1\}^N$. It is standard in the survey literature to consider a sample as a subset of U. For simplicity of notation, we define a sample i from the population U as an element of $\{0,1\}^N$. The kth coordinate of the sample i, i_k, indicates the number whether the element k is selected ($i_k = 1$) or not ($i_k = 0$). In the literature, a sample design is a function p mapping any subset of U to $[0,1]$. In this chapter, p will instead denote a probability measure p on the measurable space $\left(\mathbb{N}^N, \mathscr{P}\left(\mathbb{N}^N\right)\right)$, where $\mathscr{P}\left(\{0,1\}^N\right)$ is the power set of $\{0,1\}^N$. This measure will be called the design measure. Let I denote a random sample selected from U by means of a design measure p. I is a random variable with value in $\{0,1\}^N$ such that $I \sim \mathrm{p}$. Let π_k denote the inclusion probability of unit k, that is, the probability for unit k to be included in the sample I: we denote $\pi = (\pi_k)_{k\in\{1,\ldots,N\}}$ the vector of inclusion probabilities, when $\mathrm{E}[I] = \int i \,\mathrm{dp}(i) = \pi$. We assume that $\sum_{k\in U} \pi_k = n$, where n denotes the average sample size.
Let π_{kl} denote the probability for distinct units k and l to be jointly in the sample.

Let y denote some variable of interest. In this chapter, we are interested in estimating the population total $t_y = \sum_{k\in U} y_k$. The Horvitz-Thompson (HT) estimator is defined when $\pi \in \left]0,1\right]^N$ and is given by

$$\hat{t}_y = \sum_{k=1}^{N} y_k \frac{I_k}{\pi_k} = \sum_{k=1}^{N} d_k y_k I_k$$

where $d = (d_k)_{k\in\{1,\ldots,N\}} = (1/\pi_k)_{k\in\{1,\ldots,N\}}$ is the vector of sampling weights. This is a design-unbiased estimator for the total t_y. We look for a vector π of inclusion probabilities that minimizes, in some sense, the variance of the HT estimator. This variance is given by the so-called Horvitz-Thompson (1952) formula:

$$\mathrm{Var}\left[\hat{t}_{y\pi}\right] = \sum_{k,l\in U} \frac{y_k}{\pi_k}\frac{y_l}{\pi_l}(\pi_{kl} - \pi_k\pi_l). \tag{6.1}$$

We assume that a vector $\mathbf{x}_k = (x_{kl})_{l\in\{1,\ldots,q\}}$ of q auxiliary variables is known at the design stage for each unit k in the population. Then, $\mathbf{x}$ will denote the matrix $\left[x_{kl}\right]_{k\in\{1,\ldots,N\},l\in\{1,\ldots,q\}}$. The rows of $\mathbf{x}$ correspond to the vectors $\mathbf{x}_k$, and the columns of $\mathbf{x}$ correspond to the auxiliary variables, denoted $\mathbf{x}_{.l}$. The matrix $\mathbf{x}$ is assumed to be of full rank.

The sampling design may be improved by means of the cube method (Deville and Tillé, 2004) which enables the selection of balanced samples. The sampling design p is said to be balanced on variables $\mathbf{x}$ if

$$\mathrm{p} - a.s(i), \ \sum_{k\in U} \mathbf{x}_k \frac{i_k}{\pi_k} = t_{\mathbf{x}}, \tag{6.2}$$

where $t_{\mathbf{x}}$ gives the (vector) population total of variables $\mathbf{x}_k$. That is, the HT-estimators exactly match the known population totals. Consequently, the variance of the HT-estimator is zero for the balancing variables.

As an exact balanced sample may usually not be found, the cube method enables the selection of approximately balanced samples. The algorithm may be split into two phases, called the flight phase and the landing phase. At each step of the flight phase, one unit is either selected in the sample or definitely rejected. The result of the flight phase is given by a vector $\pi^* = (\pi_k^*)_{k\in\{1,\dots,N\}}$, where π_k^* equals 1 if unit k has been selected in the sample, 0 if unit k has been rejected from the sample, and lies between 0 and 1 otherwise. At the end of the flight phase, the balancing equations are exactly respected. That is,

$$\sum_{k\in U} \frac{\mathbf{x}_k}{\pi_k}\pi_k^* = \sum_{k\in U} \mathbf{x}_k. \tag{6.3}$$

In the case where some units are neither selected nor rejected after the flight phase, the landing phase consists in defining a sampling design among the remaining units, so that the inclusion probabilities are exactly respected and the variance due to the landing phase is minimized. Let $I = (I_k)_{k\in\{1,\dots,N\}}$ be the vector that gives the result of the landing phase.

Let $\mathrm{Cube}_{\mathbf{x},\pi}$ be the corresponding design measure, i.e, the probability law of the resulting random sample I.

Property 6.1. *If* $I \sim \mathrm{p} = \mathrm{Cube}_{\mathbf{x},\pi}$*, then*

$$\mathrm{E_p}\left[I\right] = \int_{\{0,1\}^N} i \,\mathrm{dp}(i) = \pi \tag{6.4}$$

and

$$\sum_{k\in U} \frac{\mathbf{x}_k}{\pi_k} I_k \simeq \sum_{k\in U} \mathbf{x}_k, \tag{6.5}$$

where $\mathrm{E_p}\left[\cdot\right]$ *denotes the expectation with respect to the sampling design.*

Equation (6.4) states that the inclusion probabilities are exactly respected. Equation (6.5) means that the sample is only approximately balanced, as the HT-estimator $\hat{t}_{\mathbf{x}\pi}$ usually does not exactly match the real total $t_{\mathbf{x}}$. If the sample is not exactly balanced, the sampling weights may be adjusted by means of calibration techniques (Deville and Särndal, 1992). The resulting calibration estimator of t_y is given by

$$\tilde{t}_{y,w} = \sum_{k\in U} d_k F(\lambda^{\mathrm{T}}\mathbf{x}_k) y_k I_k, \tag{6.6}$$

where $F(\cdot)$ denotes the calibration function, and λ is a vector of $\mathbb{R}^q$ that depends on $\mathbf{x}$ and I_k. A special case of (6.6) is obtained under the linear function $F(u) = 1 + u$ which leads to the generalized regression estimator

$$\hat{t}_{y,greg} = \sum_{k\in U} w_k y_k I_k, \tag{6.7}$$

where $w_k = d_k\left[1 + \left(t_{\mathbf{x}} - \hat{t}_{\mathbf{x}\pi}\right)^{\mathrm{T}}\hat{T}^{-1}\mathbf{x}_k\right]$ denotes the calibrated weight, with $\hat{T} = \sum_{k\in U} d_k \mathbf{x}_k I_k \,\mathbf{x}_k{}^{\mathrm{T}}$. Deville and Tillé (2004, section 8) give a short comparison of balanced sampling and calibration. They advocate for their joint use, since balanced sampling enables a reduction in the variability of the final weights, while calibration enables to match the known totals exactly.

A variance approximation is also provided by Deville and Tillé (2005). They suppose that the sampling design is exactly balanced, and performed with maximum entropy among sampling designs balanced on the same balancing variables, with the same inclusion probabilities. Then, under an additional assumption of

asymptotic normality of the multivariate HT-estimator under Poisson sampling, they derive the following variance approximation:

$$V_{app}(y,\pi,\mathbf{x}) = \frac{N}{N-q}\sum_{k\in U} b(\pi_k)\left(y_k - y_k^*(\pi,\mathbf{X})\right)^2, \tag{6.8}$$

where q denotes the number of balancing variables, $b(\pi_k) = 1/\pi_k - 1$ and $y_k^*(\pi) = \mathbf{x}_k\,\beta(\pi)$ is a weighted prediction of y_k obtained with the balancing variables, with

$$\beta(\pi,\mathbf{x}) = \left(\sum_{k\in U} b(\pi_k)\mathbf{x}_k\,{\mathbf{x}^{\mathrm{T}}}_k\right)^{-1}\sum_{l\in U} b(\pi_k)\mathbf{x}_k y_k$$

Other slightly different approximations are proposed in Deville and Tillé (2005), but their simulation results suggest that approximation (6.8) performs well among variance approximations that may be computed in the case of any set of inclusion probabilities.

We will retain that:

$$\begin{aligned} V_{app}(y,\pi,\mathbf{x}) &= y^{\mathrm{T}} W(\pi)\, y - y^{\mathrm{T}} W(\pi)\,\mathbf{x}\left(\mathbf{x}^{\mathrm{T}} W(\pi)\,\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}} W(\pi)\, y & (6.9)\\ &= \left\|\left((W(\pi))^{1/2}\left(\mathrm{Id}_N - \mathbf{x}\left(\mathbf{x}^{\mathrm{T}} W(\pi)\,\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}} W(\pi)\right)\right) y\right\|^2 & (6.10)\\ &= \left\|(W(\pi))^{1/2}\,\hat{\varepsilon}(y,\pi,\mathbf{x})\right\|^2 & (6.11)\end{aligned}$$

$V(y,\pi,\mathbf{x})$ is the $W(\pi)$-square norm of $\hat{\varepsilon}(y,\pi,\mathbf{x})$. The vector $\hat{\varepsilon}(y,\pi,\mathbf{x})$ is given by $\hat{\varepsilon}(y,\pi,\mathbf{x}) = \left(\mathrm{Id}_N - \mathbf{x}\left(\mathbf{x}^{\mathrm{T}} W(\pi)\,\mathbf{x}\right)^{-1}\mathbf{x}^{\mathrm{T}} W(\pi)\right) y$ and is the residual of the regression of y by $\mathbf{x}$ weighted by $W(\pi)$, where $[W(\pi)]_{k,k'} = (\pi_k)^{-1} - 1$ if $k = k'$, 0 otherwise, and $\left[(W(\pi))^{-1/2}\right]_{k,k} = \sqrt{(\pi_k)^{-1} - 1}$ if $k = k'$, 0 otherwise. The $W(\pi)$-square norm of a vector α is $\alpha^{\mathrm{T}} W(\pi)\,\alpha = \|\,(W(\pi))^{1/2}\,\alpha\|^2$.

6.3 Optimal allocation for balanced sampling

In many cases, inclusion probabilities are fixed and chosen to be proportional to an auxiliary variable known for any unit in the population at the design stage. Unequal probability sampling is then an efficient sampling design for estimating the total t_y if the variable of interest y is approximately proportional to this auxiliary variable. However, if some information on variable y is known at the design stage, it may be of interest to look for inclusion probabilities that minimize, at least approximately, the variance of the HT-estimator $\hat{t}_{y\pi}$. In what follows, section 6.3.1 mainly consists in a recall of Tillé and Favre (2005), apart from equation (6.14) which was only stated by these authors, and for which we give an explicit proof.

6.3.1 Optimal allocation for an approximation of the variance

An optimal vector π of inclusion probabilities should minimize the variance given in formula (6.1), under the constraints that

$$0 \leqslant \pi_k \leqslant 1 \text{ for any unit } k \in U \tag{6.12}$$

and

$$\sum_{k\in U}\pi_k = n. \tag{6.13}$$

Unfortunately, the variance formula (6.1) depends on second-order inclusion probabilities, and the link between the first and the second-order inclusion probabilities is intricate in case of balanced sampling, even in particular cases; see Chen et al. (1994); Deville (2000); Matei and Tillé (2005) for the special case of balanced sampling on the sample size with maximum entropy, also denominated in the literature as rejective sampling (Hájek, 1964).

Following Tillé and Favre (2005), we thus propose to minimize the variance approximation (6.8) instead. This leads to the Approximated Optimization Problem (AOP): seek for inclusion probabilities that minimize (6.8), under the constraints (6.12) and (6.13).

Property 6.2. *Let*

$$\pi = \arg\min \left\{ V_{app}(y, \pi', \mathbf{x}) \,\middle|\, \pi' \in]0,1]^N \text{ such that } \sum_{k \in U} \pi'_k = n \right\}.$$

Then

$$\forall k \in U,\ \pi_k = n \frac{|y_k - y_k^*(\pi, \mathbf{x})|}{\sum_{m \in U} |y_m - y_m^*(\pi, \mathbf{x})|}, \tag{6.14}$$

where $y_k^*(\pi, \mathbf{x}) = \mathbf{x}_k{}^{\mathrm{T}} \left(\sum_{m \in U} b_m \mathbf{x}_m \mathbf{x}^{\mathrm{T}}{}_m \right)^{-1} \sum_{m \in U} b_m \mathbf{x}_m y_m$ *and* $b_k = 1/\pi_k - 1$.

In words, the optimal inclusion probability π is proportional to the absolute value of the residuals of the regression of y on $\mathbf{x}$ weighted by $W(\pi)$.

Proof. See Appendix E.1. □

This system of equations may not be used to compute directly the optimal inclusion probabilities, since both parts of each equation depend on π. Intuitively, this formula states that if the absolute value of the residual $|e_k| = |y_k - y_k^*(\pi)|$ is large, the inclusion probability of unit k should be large, too. Conversely, a small inclusion probability should be associated with a small residual. In other words, there is no need to give large inclusion probabilities for units k such that y_k may be well predicted by the balancing variables, and attention should be paid to the remaining units instead.

A fixed-point algorithm may be used to compute the inclusion probabilities associated with formula (6.14), but the value of the variable of interest y is required for any unit in the population, and such detailed information is unknown at the design stage. This first set of inclusion probabilities is thus difficult to compute in practice.

6.3.2 Generalization of the approximated optimization problem

To overcome this difficulty, we propose a generalization of this optimization problem. Assume that a categorical variable z is known. This may be one of the balancing variables or any additional variable available at the design stage for any unit in the population. This variable defines a partition of the population into J non-overlapping subsets $U_1, \ldots, U_J$ of sizes $N_1, \ldots, N_J$, respectively, where J denotes the number of categories of the variable. Then we impose that the target inclusion probabilities satisfy the following system of equations:

$$\pi_k = \alpha_j \text{ for any unit } k \in U_j,\ j = 1, \ldots, J. \tag{6.15}$$

That is, inclusion probabilities are assumed to be equal inside each subset U_j. The variance approximation given in formula (6.8) may then be alternatively written as

$$V_{app}(y, \pi, \mathbf{x}) \equiv V(\boldsymbol{\alpha}) = \frac{N}{N-q} \sum_{j=1}^{J} b(\alpha_j) \sum_{k \in U_j} (y_k - \tilde{y}_k(\boldsymbol{\alpha}))^2, \tag{6.16}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$, $b(\alpha_j) = 1/\alpha_j - 1$ and

$$\tilde{y}_k(\boldsymbol{\alpha}) = \mathbf{x}_k{}^{\mathrm{T}} \Big(\sum_{j=1}^{J} b(\alpha_j)\mathbf{A}_j\Big)^{-1} \sum_{j=1}^{J} b(\alpha_j)\mathbf{c}_{1j}(y)$$

with $\mathbf{A}_j = \sum_{k\in U_j} \mathbf{x}_k\, \mathbf{x}_k{}^{\mathrm{T}}$ and $\mathbf{c}_{1j}(y) = \sum_{k\in U_j} \mathbf{x}_k y_k$. The General Approximated Optimization Problem (GAOP) may then be described as follows: find the $J\times 1$ vector $\boldsymbol{\alpha}$ that minimizes (6.16) under the constraints (6.12), (6.13) and (6.15). Such a vector satisfies the system of equations

$$\alpha_j = n\frac{\sigma_j(\boldsymbol{\alpha})}{\sum_{i=1}^{J} N_i\sigma_i(\boldsymbol{\alpha})}, \tag{6.17}$$

where

$$\sigma_j(\boldsymbol{\alpha}) = \frac{1}{N_j}\sum_{k\in U_j} (y_k - \tilde{y}_k(\boldsymbol{\alpha}))^2. \tag{6.18}$$

The proof is similar to that of (6.14), and is thus omitted. Note that the AOP is a special case of our setting, obtained when $J = N$. In practice, the domains associated to the variable z should be chosen so that the quantities needed for the computation of the inclusion probabilities may be either known or accurately estimated from an external source, see section 3.3.

Once again, we note that the formula (6.17) may not be directly used to compute optimal inclusion probabilities since both parts in (6.17) depend on the unknown $\boldsymbol{\alpha}$. The fixed-point Algoritm 6.1 may be used instead.

Algorithm 6.1 Fixed-point algorithm to compute optimal inclusion probabilities

Require: $\boldsymbol{\alpha}^0 = (\alpha_1^0, \ldots, \alpha_J^0)$ {Initialization}
Require: $\varepsilon \in \,]0, +\infty[$ {Threshold specification}
  $t \leftarrow 0$
  repeat
    compute $\alpha_1^{t+1}, \ldots, \alpha_J^{t+1}$ such that

$$\forall j \in \{1, \ldots, J\}\,,\ \alpha_j^{t+1} = n\frac{\sigma_j(\boldsymbol{\alpha}^t)}{\sum_{i=1}^{J} N_i\sigma_i(\boldsymbol{\alpha}^t)}.$$

    $\boldsymbol{\alpha}^{t+1} \leftarrow \left(\alpha_j^{t+1}\right)_{j\in\{1,\ldots,J\}}$
    $t \leftarrow t + 1$
  until $\max_j \|\alpha_j^t - \alpha_j^{t-1}\| < \varepsilon$
  return $\boldsymbol{\alpha}^t$

The following result states that Algoritm 6.1 always lead to a reduction in variance, as compared to the variance associated with the original $\boldsymbol{\alpha}^0$-vector.

Theorem 6.1. *At any step t of the fixed point Algoritm 6.1, $V(\boldsymbol{\alpha}^t) \leqslant V(\boldsymbol{\alpha}^{t-1})$.*

Proof. The proof is given in Appendix E.2. □

As a consequence, the sequence $(\boldsymbol{\alpha}^t)_{t\in\mathbb{N}}$ tends to a local minimum, and the approximated variance is always improved if the inclusion probabilities are given by the fixed-point algorithm. With the simulations

performed and a bound of $\epsilon = 10^{-6}$, very few iterations are required, so that $\boldsymbol{\alpha}^1$ provides a good approximation of the target vector of inclusion probabilities.

We now consider the problem of the choice of the categorical variable z whose categories are used to partition the population into domains with equal probabilities inside. Both the AOP and the GAOP should give comparable results if the absolute value of the residuals $|e_k| = |y_k - y_k^*(\pi)|$ are approximately equal inside domains $U_1, \ldots, U_J$. That is, the population U should be sorted according to the $|e_k|$ variable, and the domains separated by the fractiles of this variable. Since these residuals are practically unknown at the design-stage, an alternative consists in using the available auxiliary information. For example, qualitative variables used in the vector $\mathbf{x}_k$ of balancing variables could also be used to define the domains. Also, we previously assumed that the balancing variables did not depend on the inclusion probabilities, and in particular that no constraint on fixed size was involved in the balancing problem. This latter assumption may be relaxed if the domains inside which fixed sample size is required are used as the domains $U_1, \ldots, U_J$ in the GAOP. Let us suppose that the categorical variable z defining the domains belongs to the balancing variables. The corresponding balancing equations may be written as

$$\mathrm{p}-a.s.(i),\ \sum_{k\in U} \frac{\mathbb{1}_{U_j}(k) i_k}{\pi_k} = \sum_{k\in U} \mathbb{1}_{U_j}(k) \tag{6.19}$$

for any domain $U_j, j = 1, \ldots, J$, and the joint application of equations (6.15) and (6.19) leads to

$$\sum_{k\in U_j} I_k = \alpha_j N_j, \tag{6.20}$$

where $\sum_{k\in U_j} I_k$ denotes the size of the sub-sample $\{k \in U_j | I_k = 1\}$. The set of equations (6.20) impose that the sample size is fixed inside each domain U_j, but since $\alpha_j N_j$ may not be an integer the balancing constraints (6.20) will usually be respected to within about one unit. Note that the summation of equations (6.20) leads to

$$\begin{array}{ccccc} \sum_{k\in U} I_k & = & \sum_{j=1}^J \sum_{k\in U_j} I_k & = & \sum_{j=1}^J \alpha_j N_j \\ & = & \sum_{k\in U} \pi_k & = & n \end{array}$$

by application of equation (6.13), so that if z belongs to the balancing variables, the condition of global fixed sample size will be exactly respected.

6.3.3 Practical implementation of the optimization problem

Once again, we note that some knowledge about the variable of interest y is needed in the fixed-point algorithm. More specifically, the knowledge of $\mathbf{A}_j = \sum_{k\in U_j} \mathbf{x}_k \mathbf{x}_k{}^{\mathrm{T}}$, $\mathbf{c}_{1j}(y) = \sum_{k\in U_j} \mathbf{x}_k y_k$ and $\mathbf{c}_{2j}(y) = \sum_{k\in U_j} y_k^2$ is needed for any subset U_j. Though some of these quantities are usually unknown at the design stage, they may be replaced by estimated quantities. This is a common practice to take advantage of accurate estimated totals to improve the estimators arising from a survey, see Berger et al. (2009). For example, these estimated totals may be obtained from a previous wave or occasion of the survey, or from a larger survey; household surveys conducted in France are usually calibrated on estimates arising from the Labour Force Survey. Let us suppose that another sample I' has been selected in U with inclusion probabilities $\pi' = (\pi'_1, \ldots, \pi'_k, \ldots, \pi'_N)$. Let $\hat{\sigma}_j(\alpha)$ be obtained from (6.18) by replacing $\mathbf{A}_j$, $\mathbf{c}_{1j}(y)$ and $\mathbf{c}_{2j}(y)$ with $\hat{\mathbf{A}}'_j = \sum_{k\in U_j} \frac{\mathbf{x}_k \mathbf{x}_k{}^{\mathrm{T}}}{\pi'_k} I'_k$, $\hat{\mathbf{c}}'_{1j}(y) = \sum_{k\in U_j} \frac{\mathbf{x}_k y_k}{\pi'_k} I'_k$ and $\hat{\mathbf{c}}'_{2j}(y) = \sum_{k\in U_j} \frac{y_k^2}{\pi'_k} I'_k$. Algoritm 6.2 may then be used to compute approximately optimal inclusion probabilities, that we denote

$$\hat{\pi}^* = (\hat{\pi}^*_1, \ldots, \hat{\pi}^*_k, \ldots, \hat{\pi}^*_N) \tag{6.21}$$

where $\hat{\pi}_k^* = \hat{\alpha}_j^*$ for any unit $k \in U_j,\ j = 1, \ldots, J$. Since the exact quantities $\mathbf{A}_j$, $\mathbf{c}_{1j}(y)$ and $\mathbf{c}_{2j}(y)$ are not used in Algoritm 6.2, the computed inclusion probabilities do not necessarily lead to an optimal solution. However, the use of unbiased estimators $\hat{\mathbf{A}}_j$, $\hat{\mathbf{c}}_{1j}(y)$ and $\hat{\mathbf{c}}_{2j}(y)$ should lead to a strong reduction of the variance of the Horvitz-Thompson estimator, even if this variance is not minimized (see section 6.4.2).

Algorithm 6.2 Fixed-point algorithm to compute approximately optimal inclusion probabilities

Require: $\hat{\boldsymbol{\alpha}}^0 = (\hat{\alpha}_1^0, \ldots, \hat{\alpha}_J^0)$ {Initialization}
Require: $\varepsilon \in \,]0, +\infty[$ {Threshold specification}
  $t \leftarrow 1$
  repeat
    compute $\hat{\alpha}_1^t, \ldots, \hat{\alpha}_J^t$ such that

$$\hat{\alpha}_j^t = n \frac{\hat{\sigma}_j(\hat{\boldsymbol{\alpha}}^{t-1})}{\sum_{i=1}^J N_i \hat{\sigma}_i(\hat{\boldsymbol{\alpha}}^{t-1})} \text{ for any } j = 1, \ldots, J.$$

    $\hat{\boldsymbol{\alpha}}^t \leftarrow (\hat{\alpha}_j^t)_{j \in \{1,\ldots,J\}}$
    $t \leftarrow t + 1$
  until $\max_j \left| \hat{\alpha}_j^t - \hat{\alpha}_j^{t-1} \right| < \varepsilon$
  return $\hat{\boldsymbol{\alpha}}^* = \hat{\boldsymbol{\alpha}}^t$

We now briefly discuss the alternative solution proposed by Tillé and Favre (2005). For simplicity, we assume that the same variable of interest y and auxiliary variables $\mathbf{x}$ are collected in both the samples I' and I. First, estimated residuals

$$\hat{e}_{k1} = y_k - \mathbf{x}_k{}^{\mathrm{T}} \hat{B}' \tag{6.22}$$

are computed for units $k \in U_j$ such that $I_k = 1$, where

$$\hat{B}' = \left(\sum_{k \in U} \frac{1 - \pi'_k}{(\pi'_k)^2} \mathbf{x}_k \, \mathbf{x}_k{}^{\mathrm{T}} \, I'_k \right)^{-1} \sum_{k \in U} \frac{1 - \pi'_k}{(\pi'_k)^2} \mathbf{x}_k y_k I'_k.$$

Then, a linear model

$$|\hat{e}_{k1}|^2 = \mathbf{x}_k{}^{\mathrm{T}} \boldsymbol{\psi} + \epsilon_k \tag{6.23}$$

is postulated to predict the link between the square residuals and the auxiliary variables, where $\boldsymbol{\psi}$ is a q-vector of unknown parameters and the ϵ_ks are residuals. An estimator $\hat{\boldsymbol{\psi}}'$ of the vector $\boldsymbol{\psi}$ is obtained from sample I', to get estimated square residuals

$$|\hat{e}_{k2}|^2 = \mathbf{x}_k{}^{\mathrm{T}} \hat{\boldsymbol{\psi}}' \tag{6.24}$$

for any unit $k \in U$. Finally, the optimal inclusion probabilities are estimated by

$$\hat{\pi}_k^{TF} = n \frac{|\hat{e}_{k2}|}{\sum_{l \in U} |\hat{e}_{l2}|}. \tag{6.25}$$

If the quantities computed in (6.25) are larger than 1, Tillé and Favre (2005) propose to set the concerned inclusion probabilities to 1, while the other inclusion probabilities are recalculated. The method proposed by Tillé and Favre is less computer-intensive than the method that we propose, since no fixed-point algorithm is required for the computation of the inclusion probabilities. However, formulas (6.24) and (6.25) may lead to negative estimated square residuals and inclusion probabilities for some units in U. In that case,

the associated inclusion probabilities may be set to 0, which results in biased HT-estimators. Moreover, the quality of the prediction given in (6.24) highly depends on the predictive power of the auxiliary variables $\mathbf{x}_k$ for the residuals. If this predictive power is poor, the estimated inclusion probabilities given in (6.25) may fall far from the optimal probabilities, resulting in a possible loss of efficiency. The method proposed by Tillé and Favre (2005) as well as the method that we propose are compared in section 6.4 into a small simulation study.

6.4 A simulation study

We conducted a limited simulation study to test the performance of the procedures described in section 3. We first generated a finite population of size $N = 1\ 000$ containing 6 variables: three variables of interest $y_{.1}$, $y_{.2}$ and $y_{.3}$ and three auxiliary variables x_0, x_1 and x_2. First, the values of the variable x_0 were generated independently from a uniform distribution. The population U was divided into four groups $U_1, \ldots, U_4$ according to the quartiles of the x_0-values, and the population x_{1k} and x_{2k} were generated as

$$x_{1k} = \begin{vmatrix} 1 & \text{if } k \in U_1 \cup U_2 \\ 2 & \text{otherwise} \end{vmatrix}$$

and

$$x_{2k} = \begin{vmatrix} 1 & \text{if } k \in U_1 \cup U_3 \\ 2 & \text{otherwise.} \end{vmatrix}$$

Given the values of these auxiliary variables, the $y_{.1}$, $y_{.2}$ and $y_{.3}$-values were generated inside each group U_j according to the model

$$y_{kh} = \phi_{hj} + \eta_{kj}, h = 1, \ldots, 3. \tag{6.26}$$

The η_{jk}'s were generated according to a normal distribution with mean 0 and variance σ_j^2. The vector of model parameters $\phi_h = (\phi_{h1}, \phi_{h2}, \phi_{h3}, \phi_{h4})$ was set to $\phi_1 = (0.5, 0.5, 1.5, 1.5)$ for variable $y_{.1}$, $\phi_2 = (0.5, 1.5, 0.5, 1.5)$ for variable $y_{.2}$ and $\phi_3 = (0.2, 0.75, 1.25, 2.0)$ for variable $y_{.3}$. That is, $y_{.1}$ and $y_{.2}$ are related to the auxiliary variables x_1 and x_2 respectively, whereas the variable $y_{.3}$ is related to the interaction of variables x_1 and x_2. This last variable of interest is meant to evaluate (to some extent) the performance of the computed inclusion probabilities when the auxiliary information used is not fully adequate. We used two possible values for the vector $\sigma = (\sigma_1, \sigma_2, \sigma_3, \sigma_4)$, namely $\sigma^{(1)} = (0.2, 0.3, 0.4, 0.5)$ and $\sigma^{(2)} = (0.4, 0.6, 0.8, 1.0)$.

6.4.1 Simulation 1: optimal inclusion probabilities

We first assume that, for each variable of interest $y_{.h}$, $h = 1, \ldots, 3$, the needed quantities $\mathbf{A}_j$, $\mathbf{c}_{1j}(y_h)$ and $\mathbf{c}_{2j}(y_h)$ are exactly known. These quantities are given in Table 6.1.

The inclusion probabilities are assumed to be equal inside each group U_j. For each variable of interest y_h, $h = 1, \ldots, 3$, we note α_{hj} for the common inclusion probability for units in U_j and $\boldsymbol{\alpha}_h = (\alpha_{h1}, \alpha_{h2}, \alpha_{h3}, \alpha_{h4})'$. Algoritm 6.1 is initialized with equal probabilities $\boldsymbol{\alpha}_h^0 = (0.1, 0.1, 0.1, 0.1)$ (EQUAL). Also, two other sets of inclusion probabilities are computed: (i) probabilities $\boldsymbol{\alpha}_h^1$ obtained after the first step (FSTEP) of Algoritm 6.1 and (ii) probabilities $\boldsymbol{\alpha}_h^*$ obtained at the end (LSTEP) of Algoritm 6.1. The corresponding $\boldsymbol{\alpha}$ vectors are presented in Table 6.2. In line with formula (6.14), we note that the optimal inclusion probabilities lead to larger sample sizes in domains where the variable of interest is highly dispersed, or more precisely in domains where the balancing variables have a lower explanatory power. The values taken by the variance approximation in formula (6.8) for the three different sets of inclusion probabilities are presented in Table 6.3, as well as the totals of the variables of interest. As expected,

Table 6.1: Exact quantities needed for the computation of optimal inclusion probabilities with Algoritm 6.1, for the vectors $\sigma^{(1)}$ and $\sigma^{(2)}$

		U_1	U_2	U_3	U_4
$\mathbf{A}_j$		$\begin{pmatrix} 250 & 250 \\ 250 & 250 \end{pmatrix}$	$\begin{pmatrix} 250 & 500 \\ 500 & 1000 \end{pmatrix}$	$\begin{pmatrix} 1000 & 500 \\ 500 & 250 \end{pmatrix}$	$\begin{pmatrix} 1000 & 1000 \\ 1000 & 1000 \end{pmatrix}$
		$\sigma^{(1)}$			
$\mathbf{c}'_{1j}(\cdot)$	$y_{\cdot 1}$	$(122.16, 122.16)$	$(124.53, 249.05)$	$(752.32, 376.16)$	$(754.93, 754.93)$
	$y_{\cdot 2}$	$(129.39, 129.39)$	$(376.65, 753.31)$	$(236.70, 118.35)$	$(765.42, 765.42)$
	$y_{\cdot 3}$	$(63.75, 63.75)$	$(188.15, 376.30)$	$(620.97, 310.49)$	$(1004.64, 1004.64)$
$\mathbf{c}_{2j}(\cdot)$	$y_{\cdot 1}$	68.68	80.71	609.69	632.32
	$y_{\cdot 2}$	76.41	587.50	99.26	649.67
	$y_{\cdot 3}$	26.56	161.68	425.57	1059.42
		$\sigma^{(2)}$			
$\mathbf{c}'_{1j}(\cdot)$	$y_{\cdot 1}$	$(119.31, 119.31)$	$(124.05, 248.11)$	$(754.65, 377.32)$	$(759.85, 759.85)$
	$y_{\cdot 2}$	$(133.78, 133.78)$	$(378.31, 756.62)$	$(223.40, 111.70)$	$(780.83, 780.83)$
	$y_{\cdot 3}$	$(65.00, 65.00)$	$(188.80, 377.60)$	$(616.94, 308.47)$	$(1009.29, 1009.29)$
$\mathbf{c}_{2j}(\cdot)$	$y_{\cdot 1}$	92.92	136.29	744.27	827.01
	$y_{\cdot 2}$	109.35	652.56	222.84	864.91
	$y_{\cdot 3}$	58.11	222.91	540.47	1219.11

the approximated variance obtained with the final inclusion probabilities is systematically lower than the approximated variance obtained with the initial equal inclusion probabilities. The FSTEP and LSTEP inclusion probabilities give almost identical approximated variance, since the two sets of inclusion probabilities are very close in any case considered in the simulation, see Table 6.2. Though the results obtained after the first step may depend on the initial $\boldsymbol{\alpha}_h^0$, the vector $\boldsymbol{\alpha}_h^1$ may be expected to give a good compromise between variance reduction and low algorithmic complexity.

Table 6.2: Three sets of inclusion probabilities obtained with the fixed-point algorithm for three variables of interest, for the vectors $\sigma^{(1)}$ and $\sigma^{(2)}$

		$\sigma^{(1)}$	$\sigma^{(2)}$
$y_{\cdot 1}$	EQUAL	$(0.1, 0.1, 0.1, 0.1)$	$(0.1, 0.1, 0.1, 0.1)$
	FSTEP	$(0.055, 0.079, 0.121, 0.145)$	$(0.055, 0.079, 0.121, 0.145)$
	LSTEP	$(0.055, 0.079, 0.121, 0.145)$	$(0.055, 0.079, 0.121, 0.145)$
$y_{\cdot 2}$	EQUAL	$(0.1, 0.1, 0.1, 0.1)$	$(0.1, 0.1, 0.1, 0.1)$
	FSTEP	$(0.056, 0.081, 0.119, 0.144)$	$(0.056, 0.081, 0.119, 0.144)$
	LSTEP	$(0.056, 0.081, 0.119, 0.144)$	$(0.056, 0.081, 0.119, 0.144)$
$y_{\cdot 3}$	EQUAL	$(0.1, 0.1, 0.1, 0.1)$	$(0.1, 0.1, 0.1, 0.1)$
	FSTEP	$(0.063, 0.085, 0.119, 0.133)$	$(0.061, 0.085, 0.120, 0.134)$
	LSTEP	$(0.061, 0.085, 0.120, 0.134)$	$(0.061, 0.085, 0.120, 0.134)$

The formula (6.8) which is minimized to compute optimal inclusion probabilities only gives an approximation for the true variance, under conditions that may fail in practice. For example, Deville and Tillé

Table 6.3: Total of the variables of interest and variance approximation for three sets of inclusion probabilities

		$\sigma^{(1)}$			$\sigma^{(2)}$		
		$y_{.1}$	$y_{.2}$	$y_{.3}$	$y_{.1}$	$y_{.2}$	$y_{.3}$
Total		1 000.31	1 007.10	1 064.71	1 000.62	1 014.21	1 066.92
Variance Approximation	EQUAL	1 269.97	1 347.95	1 277.54	5 079.87	5 391.80	5 004.06
	FSTEP	1 129.58	1 189.98	1 149.92	4 518.31	4 759.92	4 494.06
	LSTEP	1 129.58	1 189.98	1 149.69	4 518.31	4 759.92	4 493.99

(2005) assume that the sampling design is exactly balanced, which is often unlikely to occur. Thus, it seems of interest to compare the performances of the different sets of inclusion probabilities with respect to the exact variance. We selected $B = 10\ 000$ samples of size $n = 100$, by balanced sampling on variables x_1 and x_2 by means of the Cube method, with the procedures EQUAL and LSTEP. Under each procedure, we computed the calibrated after balancing estimator, given by (6.6). As a measure of variability of an estimator , we used the Monte Carlo Mean Square Error (MSE) given by

$$MSE_{MC}(\hat{t}_{yw}) = \frac{1}{10\ 000}\sum_{b=1}^{B}\left(\hat{t}_{yw}(S_b) - t_y\right)^2, \tag{6.27}$$

where $\hat{t}_{yw}(I_b)$ denotes the estimator $\hat{t}_{yw}$ in the b-th sample I_b, $b = 1, \ldots, 10\ 000$. Let $\hat{t}_{yw}^{(EQUAL)}$ and $\hat{t}_{yw}^{(LSTEP)}$ denote the estimator $\hat{t}_{yw}$ under EQUAL and LSTEP, respectively. In order to compare the relative variability of the estimators, using $\hat{t}_{yw}^{(EQUAL)}$ as the reference, we used the following measure:

$$RE = \frac{MSE_{MC}(\hat{t}_{yw}^{(LSTEP)})}{MSE_{MC}(\hat{t}_{yw}^{(EQUAL)})}. \tag{6.28}$$

Table 6.4 shows the RE for the three variables. It is clear that the computed inclusion probabilities lead to a more efficient estimator in all the scenarios with a value of RE varying from 0.89 to 0.92.

Table 6.4: Relative Efficiency of the optimal vector of inclusion probabilities

	$y_{.1}$	$y_{.2}$	$y_{.3}$
$\sigma^{(1)}$	0.91	0.92	0.89
$\sigma^{(2)}$	0.89	0.90	0.92

6.4.2 Simulation 2: approximately optimal inclusion probabilities

We conducted another simulation study to take into account the practical situation when the needed quantities $\mathbf{A}_j$, $\mathbf{c}_{1j}(y_h)$ and $\mathbf{c}_{2j}(y_h)$ are unknown. That is, the computation of optimal inclusion probabilities by means of Algoritm 6.1 requires some knowledge on the variable of interest y, that may not be available at the design stage. In that case, we assume that some information has been collected on a sample I', prior to the selection of the sample I. That is, a sample I' is first selected in U, and the values of the variables

of interest y_{hk} and of the auxiliary variables $\mathbf{x}_k$ are measured for any unit $k \in U$ such that $I'_k = 1$. The needed quantities are then replaced by unbiased estimates $\hat{\mathbf{A}}'_j$, $\hat{\mathbf{c}}'_{1j}(y_h)$ and $\hat{\mathbf{c}}'_{2j}(y_h)$ (see section 6.3.3), and approximately optimal inclusion probabilities $\hat{\pi}^*_k$ given in (6.21) are obtained by means of Algoritm 6.2. The sample S is then selected by means of balanced sampling with inclusion probabilities $\hat{\pi}^*_k$. Alternatively, the method proposed by Tillé and Favre (2005) may be used instead of Algoritm 6.2 to obtain inclusion probabilities $\hat{\pi}^{TF}_k$ given by (6.25), and then to select the sample I.

We selected $B = 10,000$ samples I'_b, $b = 1, \ldots, 10\,000$ by simple random sampling of size $n_0 = 50$ (respectively, $n_0 = 100$). Then, several sets of inclusion probabilities are computed for any unit $k \in U$. The inclusion probabilities are equal inside each group U_j. For each sample I'_b, Algoritm 6.2 is initialized with equal probabilities $\hat{\boldsymbol{\alpha}}^0_h = (0.1, 0.1, 0.1, 0.1)$ (EQUAL). Two other sets of inclusion probabilities are computed: (i) probabilities $\hat{\boldsymbol{\alpha}}^*_{bh}$ obtained at the end (APPROX) of Algoritm 6.2, and (ii) probabilities $\hat{\boldsymbol{\alpha}}^{TF}_{bh}$ associated to the method of Tillé and Favre (MODEL). Then, a sample I_b, $b = 1, \ldots, 10\,000$ of size $n = 100$ is selected by balanced sampling on variables x_1 and x_2 by means of the Cube method, with the procedures EQUAL, APPROX and MODEL.

To compare the approximately optimal inclusion probabilities associated to the procedures APPROX and MODEL with the true, optimal inclusion probabilities associated to the LSTEP procedure (see section 6.4.1), we used the Monte Carlo Mean (MEAN), given by

$$MEAN_{MC}\left(\hat{\boldsymbol{\alpha}}^{(.)}_h\right) = \frac{1}{10\,000}\sum_{b=1}^{B}\hat{\boldsymbol{\alpha}}^{(.)}_{bh}. \tag{6.29}$$

We present in Table 6.5 the Monte Carlo Mean obtained with APPROX and MODEL and a size of $n_0 = 50$ for the prior sample. The results obtained with $n_0 = 100$ were almost identical, and are thus omitted. Clearly, the Monte Carlo Bias associated to the proposed method is negligible so that APPROX may be expected to give results close to that of LSTEP. On the other hand, the Monte Carlo Bias associated to MODEL is non-negligible, except for the variable $y_{.3}$, which may result in a loss of efficiency. To evaluate the performances of each procedure, we computed for each of them the calibrated after balancing estimator, given by (6.6). As a measure of variability of an estimator , we used the Monte Carlo Mean Square Error (MSE) given by equation (6.27) where $\hat{t}_{yw}(S_b)$ denotes the estimator $\hat{t}_{yw}$ in the b-th sample S_b, $b = 1, \ldots, 10\,000$. Let $\hat{t}^{(EQUAL)}_{yw}$, $\hat{t}^{(APPROX)}_{yw}$ and $\hat{t}^{(MODEL)}_{yw}$ denote the estimator $\hat{t}_{yw}$ under EQUAL, APPROX and MODEL, respectively. In order to compare the relative variability of the estimators, using $\hat{t}^{(EQUAL)}_{yw}$ as the reference, we used the following measure:

$$RE = \frac{MSE_{MC}(\hat{t}^{(.)}_{yw})}{MSE_{MC}(\hat{t}^{(EQUAL)}_{yw})}. \tag{6.30}$$

The results are presented in Table 6.6. Once again, we note that the inclusion probabilities computed with APPROX lead to a more efficient estimator than EQUAL, with values of RE ranging from 0.88 to 0.95. We note that the RE is closer to one when the sample size decreases. That is, a smaller size of the external survey used to estimate the needed quantities results in a loss of accuracy of the computed inclusion probabilities, as could be expected. Therefore, we advocate for the use of domains in which these needed quantities may be precisely estimated. Also, we note that MODEL gives quite poor results since it is outperformed by APPROX in all cases, and by EQUAL in 10 out of 12 cases.

Table 6.5: Comparison of Algoritm 6.1 Tillé and Favre method

This table contains the optimal inclusion probabilities given by Algoritm 6.1 and Monte Carlo Mean of the approximately optimal inclusion probabilities given by Algoritm 6.2 or by the method of Tillé and Favre for three variables of interest, obtained with $n_0 = 50$ for the vectors $\sigma^{(1)}$ and $\sigma^{(2)}$.

		$\sigma^{(1)}$	$\sigma^{(2)}$
$y_{.1}$	LSTEP	$(0.055, 0.079, 0.121, 0.145)$	$(0.055, 0.079, 0.121, 0.145)$
	APPROX	$(0.055, 0.079, 0.121, 0.144)$	$(0.055, 0.079, 0.121, 0.144)$
	MODEL	$(0.047, 0.084, 0.126, 0.143)$	$(0.047, 0.084, 0.126, 0.143)$
$y_{.2}$	LSTEP	$(0.056, 0.081, 0.119, 0.144)$	$(0.056, 0.081, 0.119, 0.144)$
	APPROX	$(0.056, 0.081, 0.119, 0.144)$	$(0.056, 0.081, 0.119, 0.144)$
	MODEL	$(0.047, 0.086, 0.124, 0.143)$	$(0.047, 0.086, 0.124, 0.143)$
$y_{.3}$	LSTEP	$(0.061, 0.085, 0.120, 0.134)$	$(0.061, 0.085, 0.120, 0.134)$
	APPROX	$(0.061, 0.085, 0.121, 0.134)$	$(0.061, 0.085, 0.120, 0.134)$
	MODEL	$(0.061, 0.086, 0.120, 0.133)$	$(0.060, 0.085, 0.121, 0.134)$

Table 6.6: Relative Efficiency when prior information is available

We compare the relative efficiency for two vectors of inclusion probabilities computed with respect to prior information known from a past survey.

		$n_0 = 50$			$n_0 = 100$		
		$y_{.1}$	$y_{.2}$	$y_{.3}$	$y_{.1}$	$y_{.2}$	$y_{.3}$
$\sigma^{(1)}$	APPROX	0.93	0.95	0.95	0.88	0.91	0.89
	MODEL	1.13	1.20	1.31	1.00	1.04	0.98
$\sigma^{(2)}$	APPROX	0.93	0.89	0.94	0.92	0.88	0.90
	MODEL	1.13	1.22	1.17	1.03	1.01	0.96

6.5 Optimal inclusion probabilities for probabilistic quota sampling

6.5.1 Probabilistic quota sampling

Suppose those responsible for the survey wish to sample a certain number of people, respecting marginal constraints on age categories and gender, like in quota sampling. For all element of the population, the probability to be selected is known and controlled.

We assume that the population U is divided in $J \times L$ groups, $J, L \in \mathbb{N}$. A group is denoted $U_{j,l}$ and $U = \bigcup_{(j,l)\in\{1,\dots,J\}\times\{1,\dots,J\}} U_{jl}$. For $j \in \{1,\dots,J\}$, $l \in \{1,\dots,L\}$, let $U_{j.} = \bigcup_l U_{jl}$, $U_{.l} = \bigcup_j U_{jl}$, . $N_{j.} = \#(U_{j.})$, $N_{.l} = \#(U_{.l})$ and $N_{jl} = \#(U_{jl})$

Define the variables $(\mathbf{x}_{.,l})_{l\in\{1,\dots,J+L-1\}}$: for $l \in \{1,\dots,J\}$, $x_{.l} = (x_{kl})_{k\in\{1,\dots,N\}}$, $x_{.,j} = (x_{kj})_{k\in\{1,\dots,N\}}$, with

$$x_{k,l} = \begin{vmatrix} 1 & \text{if } k \in U_{j.} \\ 0 & \text{otherwise,} \end{vmatrix}$$

and for $l \in \{J+1,\dots,J+L-1\}$,

$$x_{k,J+l} = \begin{vmatrix} 1 & \text{if } k \in U_{.l} \\ 0 & \text{otherwise.} \end{vmatrix}$$

6.5.2 Optimal inclusion probabilities

We want to determine optimal inclusion probabilities, constant on the sub-populations U_{ij}, that minimize the approximate variance of a sample balanced on a number of elements selected within each sub-population $U_{j,\cdot}$ equal to $n_{j.}$ and on a number of elements selected within each sub-population $U_{\cdot,j}$ equal to $n_{.l}$, where $n_{j.}, n_{.l} \in \mathbb{N}$, $\sum_{j=1}^{J} n_{j.} = \sum_{l=1}^{L} n_{.l} = n$. This is approximately obtained with $\mathrm{p} = \mathrm{Cube}_{\pi,\mathbf{x}*\pi}$, where $\mathbf{x} * \pi$ is the matrix $\mathbf{x} * \pi = \left[\pi_k x_{kl}\right]_{k\in\{1,\dots,N\},l\in\{1,\dots,J+L-1\}}$.

We look for the inclusion probabilities that minimize the variance of such a design, and that is constant on each subpopulation U_{jl}, that is the vector:

$$\pi^* = \arg\min \left\{ V_{app}\left(y, \pi, \mathbf{x} * \pi\right) | \, \pi \in \left]0,1\right]^N A\pi = a \right\}, \tag{6.31}$$

where A is a matrix, a is a vector such that

$$[A\pi = a] \Leftrightarrow [\forall j \in \{1,\dots,J\}, \forall l \in \{1,\dots,L\}, k, k' \in U_{jl}, \pi_k = \pi_{k'} \text{ and } \sum_{k\in U} \pi_k = n].$$

In that case, the fixed point method cannot be applied, because $\mathbf{x} * \pi$ depends on π. The function $\pi \mapsto V_{app}(y, \pi, \mathbf{x} * \pi)$ can be defined by continuity on $[0,1]^N$, and is continuous. There exists a global miminum to this function. To compute optimal inclusion probabilities, we will use an iterative method that allows the computation of a local minimum of $\pi \mapsto V_{app}(y, \pi, \mathbf{x} * y)$, in a more general case. We define

$$V : \pi \mapsto V_{app}(y, \pi, \mathbf{x} * \pi),$$

$$
\begin{aligned}
\frac{\mathrm{d}V}{\mathrm{d}\pi}(\pi): \quad \mathbb{R}^p &\to \mathbb{R} \\
\delta &\mapsto y^{\mathrm{T}}\left(\left(\frac{\mathrm{d}W}{\mathrm{d}\pi}(\pi)\right)(\delta)\right)y \\
&\quad -\left(\frac{\mathrm{d}}{\mathrm{d}\pi}\left(y^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\left((\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\right)^{-1}(\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)y\right)(\pi)\right)(\delta) \\
&= y^{\mathrm{T}}\left(\left(\frac{\mathrm{d}W}{\mathrm{d}\pi}(\pi)\right)(\delta)\right)y \\
&\quad -2\,y^{\mathrm{T}}\left(\left(\frac{\mathrm{d}W}{\mathrm{d}\pi}(\pi)\right)(\delta)\right)(\mathbf{x}*\pi)\left((\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\right)^{-1}(\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)y \\
&\quad -2\,y^{\mathrm{T}}W(\pi)(\mathbf{x}*\delta)\left((\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\right)^{-1}(\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)y \\
&\quad +y^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\left((\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\right)^{-1} \\
&\qquad \Big((\mathbf{x}*\delta)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi) \\
&\qquad\quad +(\mathbf{x}*\pi)^{\mathrm{T}}\left(\left(\frac{\mathrm{d}W}{\mathrm{d}\pi}(\pi)\right)(\delta)\right)(\mathbf{x}*\pi) \\
&\qquad\quad +(\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\delta)\Big) \\
&\qquad \left((\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)(\mathbf{x}*\pi)\right)^{-1}(\mathbf{x}*\pi)^{\mathrm{T}}W(\pi)y
\end{aligned}
$$

and

$$
(\mathrm{grad}(V))(\pi) = \sum_k \left(\frac{dV}{d\pi}(\pi)\right)(e_k).e_k,
$$

where e_k is the vector $(e_{k1}\dots e_{kN})$, with $e_{kl} = 1$ if $k = l$, 0 otherwise. To get a local minimum, we use a method of gradient descent (see Snyman (2005)):

Algorithm 6.3 Gradient descent to determine local optimal inclusion probabilities

Require: $\pi^0 \in]0,1]^N$ {Initialization}
Require: $\varepsilon \in]0,+\infty[$ {Threshold specification}
  $\pi^1 \leftarrow \left(I - \left(A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\right)\right)\pi^{(0)}$ {Projection of π^0 in the constrained space}
  $t \leftarrow 1$
  repeat
    $\Delta^t \leftarrow \left(I - \left(A\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\right)\right)(-\mathrm{grad}(V))(\pi)$ {Descent direction calculation}
    $\mathrm{step}^t \leftarrow \min\left\{k \in \mathbb{N} \middle| \pi^t + \frac{\Delta^t}{k} \in]0,1]^N, V\left(\pi^t + \frac{\Delta^t}{k}\right) < V\left(\pi^t\right)\right\}$ {Step calculation}
    $\pi^{t+1} \leftarrow \pi^t + \frac{\Delta^t}{\mathrm{step}^t}$ {Calculation of next iteration}
    $t \leftarrow t+1$
  until $\left\|\pi^t - \pi^{t-1}\right\|_\infty < \varepsilon$
  return π^t

Like in section 6.4, it is possible to show that the knowledge of $\sum_{k\in U_{jl}} y_k$ and $\sum_{k\in U_{jl}} y_k^2$ for all $j \in \{1,\dots,J\}, l \in \{1,\dots,L\}$ is sufficient for the calculation of $\frac{\mathrm{d}V}{\mathrm{d}\pi}(\pi)$ when $A\pi = a$.

If the quantities $\sum_{k\in U_{jl}} y_k$ and $\sum_{k\in U_{jl}} y_k^2$ for all $j \in \{1,\dots,J\}, l \in \{1,\dots,L\}$ can be estimated, it is possible to run the Algorithm 6.3 by using the estimates instead of the true values. As V is a continuous

function of these quantities, one can expect that the use of inclusion probabilities that will result from the algrithm will lead to a smaller variance.

6.6 Concluding remarks

In this chapter, we studied the problem of computation of inclusion probabilities in the context of balanced sampling. We showed that, under some conditions on the vector of balancing variables, the computation earlier suggested by Tillé and Favre (2005) to obtain inclusion probabilities systematically leads to a decrease of the variance of the Horvitz-Thompson estimator. This algorithm requires that some quantities may be known from an external source. If not, we proposed an alternative algorithm where the needed quantities are estimated. This situation is not uncommon in practice; since most surveys are periodic, it may be of interest to take advantage of the previous waves of a survey. Results from a limited simulation study have shown that, even in the latter case, a significant decrease of the variance may be expected.
Further investigation on the matter is needed. First, the case where the balancing variables include some fixed-size constraints on domains is not covered by the Tillé and Favre (2005) algorithm, if these domains do not coincide with those used in the GAOP. Such constraints are frequently needed, for example if a given level of precision is required for certain subdivisions of the population. We give an example where the Tillé and Favre (2005) algorithm does not apply, the probabilistic quota sampling, and then we propose a second algorithm that includes all kind of linear constraints on the inclusion probabilities and all balancing variables that may depend on π. This algorithm just allows to get local optimum inclusion probabilities. Secondly, the approximation of variance of Deville and Tillé (2005) used to compute the inclusion probabilities is unlikely to hold if the assumption of maximum entropy is not satisfied.

References

Berger, Y. G., Muñoz, J. F., and Rancourt, E. (2009). Variance estimation of survey estimates calibrated on estimated control totals– An application to the extended regression estimator and the regression composite estimator. *Computational Statistics & Data Analysis*, 53(7):2596 – 2604.

Chen, X., Dempster, A., and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3):457.

Deville, J.-C. (2000). Note sur l'algorithme de chen, dempster et liu. *Rapport technique, CREST-ENSAI, Rennes*.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912.

Deville, J.-C. and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128(2):569–591.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35(4):1491–1523.

Matei, A. and Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21(4):543.

Snyman, J. (2005). *Practical Mathematical Optimization: an Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, volume 97. Springer Verlag.

Tillé, Y. and Favre, A. (2005). Optimal allocation in balanced sampling. *Statistics & Probability Letters*, 74(1):31–37.

Conclusion

If the analyst considers design variables as random, and postulates a model on the design-variable distribution given the response, then, under weak conditions on the informative selection (weak dependence of draws given the responses values), it is possible to consider sample responses as iid $\rho_\infty f.\lambda$, where $\rho_\infty(y)$ is a weighted function, defined as the limit of the expectation of the number of times an element will be selected given its associated response value equals y, divided by the global rate of selection. Under mild assumptions, the empirical sample cdf converges to the weighted version of the population cdf, and kernel density estimators converge to the sample pdf. These results may allow to make inference on the population pdf, and are indications that the sample responses may behave asymptotically as if they were iid $(\rho_\infty f.\lambda)$. In addition, it offers the opportunity to replace the true sample likelihood by an approximate likelihood, derived as the product of limit sample pdf's. This is also an alternative to Horvitz-Thompson plug-in methods: when the selection is weakly informative, and when the inclusion probabilities are highly variable, then the inference based on the weighted distribution can prove more efficient.

Concerning the computation of optimal inclusion probabilities for balanced sampling, we propose an algorithm that should lead to a reduced variance of the estimator of the total of a study variable when the sample design is with maximum entropy and balanced on two qualitative design variables y, and those responsible for the survey have a prior on the dispersion of y in the corresponding sub-populations.

Appendix A

Essential mathematical concepts and notation used in the dissertation

A.1 Set theoretic notation and terminology

The symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ denote the sets of natural integers, relative integers, rational numbers and real numbers. Given $(a,b) \in \mathbb{R}^2$, we define the intervals $[a,b] = \{x \in \mathbb{R} \mid a \leqslant x \leqslant b\}$, $]a,b] = \{x \in \mathbb{R} \mid a < x \leqslant b\}$, $[a,b[= \{x \in \mathbb{R} \mid a \leqslant x < b\}$ and $]a,b[= \{x \in \mathbb{R} \mid a < x < b\}$. Given a set A, $\mathscr{P}(A)$ denotes the power set of A, i.e. set of subsets of A, $\#(A)$ is the cardinal number of A.

A function from a set A to a set B will often be defined by using the notation:

$$f : A \to B, a \mapsto f(a).$$

The set of functions from a set A to a set B is denoted B^A or $\mathscr{F}(A,B)$. The image set of a function f from a set A to a set B is denoted $f(A)$ or $\mathrm{Im}(f)$.

Given A a subset of a set E, $\mathbb{1}_A$ is the function: $\mathbb{1}_A : E \to \{0,1\}, a \mapsto 1$ if $a \in A$, 0 otherwise. For $N \in \mathbb{N}\backslash\{0\}$, $\mathfrak{S}_N$ is the set of permutations of $\{1,\ldots,N\}$. The cardinal number of $\mathfrak{S}_N$ is factorial N and is denoted $N!$.

For $N \in \mathbb{N}\backslash\{0\}$, $n \in \mathbb{N}$, for $r \in \mathscr{F}(\{1,\ldots,N\},\{1,\ldots,n\})$, $a \in A^N$, $r.a$ denotes the vector of A^n $r.a = \left(a_{r(1)} \ldots a_{r(n)}\right)$. Given two sets A, B, with $\#A < \#B$, $\mathrm{Inj}(A,B)$ is the set of injective functions from A to B. The number $\mathrm{Inj}(\{1,\ldots,n\},\{1,\ldots,N\})$ is denoted $\binom{N}{n}$.

A sequence of elements of a set A is an element $a = (a_\gamma)_{\gamma\in\mathbb{N}} \in A^{\mathbb{N}}$. Given a sequence on $\mathbb{R}^d$,

A.2 Derivation

The derivative of order l of a real function f is denoted $f^{(l)}$. For $l \in \mathbb{N}$, the set of l differentiable functions is denoted $\mathscr{C}^{(l)}$.

A.3 Measure and probability

Given a set $\mathcal{A}$, and $a \in \mathcal{A}$, δ_a is the Dirac measure defined on the measurable space $(\mathcal{A}, \mathscr{P}(\mathcal{A}))$ by:

$$\delta_a : \mathscr{P}(A) \to \{0,1\} : A \mapsto \begin{vmatrix} 1 & \text{if } a \in A \\ 0 & \text{otherwise,} \end{vmatrix}$$

and for $A \in \mathscr{P}(\mathcal{A})$, δ_A is the counting measure on A defined on the measurable space $(\mathcal{A}, \mathscr{P}(\mathcal{A}))$ by $\delta_A = \sum_{a\in A} \delta_a$. The uniform probability measure on a finite set A is $\#(A)^{-1}\delta_A$.

Given a random variable Y defined on the measure space $(\Omega, \mathscr{A}, \mathrm{P})$, with value in a measurable space, P^Y is the probability distribution of Y, that is the probability measure on $(\mathscr{Y}, \mathscr{T}_Y)$:

$$\mathrm{P}^Y : \mathscr{T}_Y \to [0,1], A \mapsto \mathrm{P}\left(Y^{-1}(A)\right).$$

Given a topological space E with a topology, $\mathscr{B}_E$ denotes the smallest σ-algebra that contains the open sets of E.

A Radon measure on $\mathscr{B}_{\mathbb{R}^d}$ is a measure which is finite on all the compact sets of $\mathscr{B}_{\mathbb{R}^d}$

Given a probability measure μ, $Y \sim \mu \to \mathrm{P}^Y = \mu$. Given a density on the measure space $(\Omega, \mathscr{A}, \mu)$, that is a measurable positive function with value in $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\int f d\mu = 1$, $f.\mu$ is the probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ defined by $f.\mu : A \mapsto \int \mathbb{1}_A f \,\mathrm{d}\mu$. A positive Radon measure μ on a Polish topological space is a positive measure on the Borel σ-algebra, such that the measure of all compact sets is finite.

Let A a countable set, $(Y_a)_{a\in A}$ a collection of random variables defined on a probability space $(\Omega, \mathscr{A}, \mathrm{P})$ then $(Y_a)_{a\in A}$ is exchangeable if $\forall r \in \mathbb{N}$ such that $r < \#A$, $k_1 \dots k_r \in A$ distinct, $l_1 \dots l_r \in A$ distinct,

$$P^{\left((Y_{k_1})\dots(Y_{k_r})\right)} = P^{\left((Y_{l_1})\dots(Y_{l_r})\right)} \tag{A.1}$$

The Lebesgue measure on $\mathbb{R}$ is denoted λ. For $p \in \mathbb{N}$, the Lebesgue measure on $\mathbb{R}^p$ is denoted λ_p or $\lambda_{\mathbb{R}^p}$. Given a Borel set B, the Lebesgue measure on B is denoted λ_B.

Some probability laws The normal distribution with mean parameter β, and variance parameter σ^2 is denoted $\mathscr{N}(\beta, \sigma^2)$. The cdf of $\mathscr{N}(0,1)$ is denoted Φ, the density with respect to the Lebesgue measure is denoted ϕ, and the quantile function Φ^{-1}. The Gamma distribution of parameters with shape parameter k, and scale parameter θ, is denoted $\mathrm{Gamma}(k, \theta)$. The associated pdf is the function: $y \mapsto y^{k-1}\left(\Gamma(k)\,\theta^k\right)^{-1}\exp\left(-y/\theta\right)$, where Γ is the Gamma function: $\Gamma : \mathbb{R}\backslash\mathbb{Z}^- \to \mathbb{R}, z \mapsto \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}t$. Its cdf is the function: $y \mapsto (\Gamma(k))^{-1}(\gamma(k, y/\theta))$, with $\gamma(s,x) = \int_0^x t^{s-1}\,e^{-t}\,\mathrm{d}t$.

Stochastic o and O Let $(X_\gamma)_{\gamma\in\mathbb{N}}$ a sequence of r.v. $(\mathbb{R}^m, \mathscr{B}_{R^m})$ and f a positive function on $\mathbb{N}$. Then $X_\gamma = o_{\mathrm{P}}(f(\gamma))$ denotes

$$\forall \varepsilon \in \mathbb{R}^+, \lim_\gamma \mathrm{P}\left(\|X_\gamma\| > f(\gamma)\varepsilon\right) = 0,$$

and $X_\gamma = O_{\mathrm{P}}(f(\gamma))$ denotes

$$\forall \varepsilon > 0, \exists M_\varepsilon > 0, \Gamma_\varepsilon \in \mathbb{N} \text{ such that } \forall \gamma > \Gamma_\varepsilon, \mathrm{P}\,\|X_\gamma\| > M_\varepsilon f(\gamma) < \varepsilon.$$

A.4 Algebra

For $N \in \mathbb{N}$, the identity matrix of order N, is denoted Id_N. The image of a matrix M by a transpose operator is denoted M^{T}.

A.5 Miscellaneous

We define the function: $\lfloor . \rfloor : \mathbb{R} \to \mathbb{N}, x \mapsto \lfloor x \rfloor = \max\{y \in \mathbb{N} | y \leqslant x\}$.

Appendix B

Proofs for chapter 2

B.1 Proof of Property 2.7

Proof. Let $A \in \mathscr{B}_p$. The sample size n_γ is not random, and takes the value $n_\gamma^* \in \mathbb{N}$. There exists a relation between n_γ^*, N_γ and $\int m_\gamma f_\gamma \, \mathrm{d}\mu_Y$:

$$n_\gamma^* \quad = \quad \mathrm{E}\left[n_\gamma\right] = \mathrm{E}\left[\sum_{k=1}^{N_\gamma} I_{\gamma k}\right] = N_\gamma \int m_\gamma f_\gamma \, \mathrm{d}\mu_Y .$$

Letting $\ell \in \left\{1, \ldots, n_\gamma^*\right\}$, we calculate:

$$
\begin{aligned}
\mathrm{P}_{\theta,\xi}\left(\left\{Y_{\gamma R_\gamma(\ell)} \in A\right\}\right) &= \sum_{k=1}^{N_\gamma} \mathrm{P}\left(\{Y_{\gamma k} \in A\} \cap \{R_\gamma(\ell) = k\}\right) \\
&= \sum_{k=1}^{N_\gamma} \mathrm{P}\left(\{Y_{\gamma k} \in A\} \cap \{R_\gamma(\ell) = k\}\right) \\
&= N_\gamma \, \mathrm{P}\left(\{Y_{\gamma 1} \in A\} \cap \{R_\gamma(\ell) = 1\}\right) \\
&= N_\gamma \sum_{i=0}^{n_\gamma^*} \mathrm{P}\left(\{Y_{\gamma 1} \in A\} \cap \{I_{\gamma 1} = i\} \cap \{R_\gamma(\ell) = 1\}\right) \\
&= N_\gamma \sum_{i=0}^{n_\gamma} \mathrm{P}\left(\{Y_{\gamma 1} \in A\}\right) \mathrm{P}\left(I_{\gamma 1} = i \,|Y_{\gamma 1} \in A\right) \mathrm{P}\left(R_\gamma(\ell) = 1 | \{Y_{\gamma 1} \in A\} \cap \{I_{\gamma 1} = i\}\right) \\
&= N_\gamma \sum_{i=0}^{n_\gamma} \mathrm{P}\left(\{Y_{\gamma 1} \in A\}\right) \mathrm{P}\left(I_{\gamma 1} = i \,|Y_{\gamma 1} \in A\right) \frac{i}{n_\gamma^*} \\
&= \frac{N_\gamma}{n_\gamma^*} \mathrm{P}\left(\{Y_{\gamma 1} \in A\}\right) \mathrm{E}\left[I_{\gamma 1} \,|Y_{\gamma 1} \in A\right] \\
&= \frac{N_\gamma}{n_\gamma^*} \mathrm{P}\left(\{Y_{\gamma 1} \in A\}\right) \frac{\int_A \mathrm{E}\left[I_{\gamma 1} \,|Y_{\gamma 1} = y\right] \mathrm{dP}^{Y_{\gamma 1}}(y)}{\mathrm{P}\left(\{Y_{\gamma 1} \in A\}\right)} \\
&= \frac{\int_A \mathrm{E}\left[I_{\gamma k} | Y_{\gamma k} = y\right] f_\gamma(y) \, \mathrm{d}\mu_Y(y)}{\int m_\gamma f_\gamma \, \mathrm{d}\mu_Y} \\
&= \int_A \frac{m_\gamma}{\int m_\gamma f_\gamma \, \mathrm{d}\mu_Y} f_\gamma \, \mathrm{d}\mu_Y \\
&= \int_A \rho_\gamma f_\gamma \, \mathrm{d}\mu_Y .
\end{aligned}
$$

Then,

$$
\mathrm{P}_{\theta,\xi}^{Y_{\gamma R_\gamma(\ell)}} = \left(\rho_\gamma f_\gamma\right) . \mu_Y .
$$

□

Appendix C

Proofs for chapter 3

C.1 Proofs of Theorems 3.1 and 3.2

The first subsection contains the proof of Theorem 3.1. The proof consists in showing the uniform L_2 convergence of the empirical cdf, seen as a ratio of two random variables. First, we show that from A3.3a we can deduce the L_2 convergence of both the numerator and denominator, then the classical proof of Glivenko-Cantelli is adapted to obtain a uniform L_2 convergence.

The second subsection contains the proof of Theorem 3.2. We first construct two sequences of random variables $(\mathcal{I}'_\gamma)$ and Y' such that $\forall\gamma$, $(\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma)$ and $(\mathcal{I}_\gamma, \mathcal{Y}_\gamma)$ have the same distribution. We then prove uniform L_2 convergence of the empirical cdf defined from $(\mathcal{I}'_\gamma)$ and Y', almost surely in Y'. The result is "design-based" in the sense that it is conditional on Y', and is of independent interest. We conclude by showing the almost sure convergence.

C.1.1 Proof of Theorem 3.1: uniform L_2 convergence of the empirical cdf

Lemma C.1. *Given a bounded measurable function $b : \mathbb{R} \to \mathbb{R}$, A3.0 and A3.3a imply that*

$$\frac{\sum_{k\in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma} \underset{\gamma\to\infty}{\overset{L_2}{\to}} \int bmf \, d\lambda.$$

Proof. Assume A3.0 and A3.3a. The exchangeability property (2.9) implies

$$\mathrm{E}\left[\frac{\sum_{k\in U_\gamma} b(Y_k) I_{\gamma k}}{N_\gamma}\right] = \frac{\sum_{k\in U_\gamma} \mathrm{E}\left[b(Y_k) I_{\gamma k}\right]}{N_\gamma} = \int bm_\gamma f \, d\lambda \underset{\gamma\to\infty}{\to} \int bmf \, d\lambda$$

by A3.0a, A3.0b and the dominated convergence theorem. Further, (2.9) implies

$$\begin{aligned}
&\operatorname{Var}\left[\frac{\sum_{k\in U_\gamma} b(Y_k)I_{\gamma k}}{N_\gamma}\right] \\
&= \frac{1}{N_\gamma^2}\sum_{k,\ell\in U_\gamma}\{\operatorname{Cov}\left(b(Y_k)\,\mathrm{E}\left[I_{\gamma k}\mid Y_k,Y_\ell\right],b(Y_\ell)\,\mathrm{E}\left[I_{\gamma l}\mid Y_k,Y_\ell\right]\right) \\
&\qquad + \mathrm{E}\left[b(Y_k)b(Y_\ell)\operatorname{Cov}\left(I_{\gamma k},I_{\gamma\ell}\mid Y_k,Y_\ell\right)\right]\} \\
&= \left(1-\frac{1}{N_\gamma}\right)\left(\int b(y_1)b(y_2)m'_\gamma(y_1,y_2)m'_\gamma(y_2,y_1)f(y_1)f(y_2)dy_1dy_2\right. \\
&\qquad\qquad -\left(\int b(y_1)m'_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right)^2 \\
&\qquad\qquad \left.+\int b(y_1)b(y_2)c_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right) \\
&\quad +\frac{1}{N_\gamma}\left(\int b^2v_\gamma f\,d\lambda+\int b^2m_\gamma^2 f\,d\lambda-\left(\int bm_\gamma f\,d\lambda\right)^2\right) \\
&= \left(1-\frac{1}{N_\gamma}\right)\left(\int b(y_1)b(y_2)\left(m'_\gamma(y_1,y_2)m'_\gamma(y_2,y_1)\right.\right. \\
&\qquad\qquad \left.-m_\gamma(y_1)m_\gamma(y_2)\right)f(y_1)f(y_2)dy_1dy_2 \\
&\qquad\qquad \left.+\int b(y_1)b(y_2)c_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2\right) \\
&\quad +\frac{1}{N_\gamma}\left(\int b^2(v_\gamma+m_\gamma^2)f\,d\lambda-\left(\int bm_\gamma f\,d\lambda\right)^2\right) \\
&= o_\gamma(1)
\end{aligned}$$

by A3.3a, A3.3b, and A3.3c, and the result is proved. □

Lemma C.2. *Under A3.0 and A3.3a, the numerator of the empirical cdf converges uniformly in L_2:*

$$\lim_{\gamma\to\infty}\mathrm{E}\left[\left(\sup_{\alpha\in\mathbb{R}}\left|\frac{\sum_{k\in U_\gamma}\mathbb{1}_{(-\infty,\alpha]}(Y_k)I_{\gamma k}}{N_\gamma}-\int\mathbb{1}_{(-\infty,\alpha]}m_\gamma f\,d\lambda\right|\right)^2\right]=0.$$

Proof. We first define $G_\gamma:\mathbb{R}\to\mathbb{R}^+$ and $G_s:\mathbb{R}\to\mathbb{R}^+$ as

$$G_\gamma(\alpha)=\frac{1}{N_\gamma}\sum_{k\in U_\gamma}\mathbb{1}_{(-\infty,\alpha]}(Y_k)I_{\gamma k}\text{ and }G_s(\alpha)=\int\mathbb{1}_{(-\infty,\alpha]}mf\,d\lambda.$$

With these definitions,

$$\sup_{\alpha\in\mathbb{R}}\left|\frac{\sum_{k\in U_\gamma}\mathbb{1}_{(-\infty,\alpha]}(Y_k)I_{\gamma k}}{N_\gamma}-\int\mathbb{1}_{(-\infty,\alpha]}m_\gamma f\,d\lambda\right|=\|G_\gamma-G_s\|_\infty.$$

Let $\eta\in\mathbb{N}^*$ index the positive integers and define a sequence of subdivisions $\{\alpha_{\eta,q}\}_{q=0}^{\eta+1}$ of $\mathbb{R}$ via $\alpha_{\eta,0}=-\infty$, $\alpha_{\eta,\eta+1}=\infty$, and for $q=1,\ldots,\eta$,

$$\alpha_{\eta,q}=\inf\left\{\alpha\in\mathbb{R}\,\middle|\,G_s(\alpha)\geqslant\eta^{-1}qG_s(\infty)\right\}.$$

We first show that for all positive integers η,

$$\sup_{\alpha\in\mathbb{R}}\{|G_\gamma(\alpha)-G_s(\alpha)|\}\leqslant\max_{0\leqslant q\leqslant\eta+1}\{|G_\gamma(\alpha_{\eta,q})-G_s(\alpha_{\eta,q})|\}+\frac{G_s(\infty)}{\eta}.$$

Let $\eta\in\mathbb{N}$ and $\alpha\in\mathbb{R}$. Then $\alpha\in[\alpha_{\eta,q},\alpha_{\eta,q+1}]$ for some $0\leqslant q\leqslant\eta$, and

$$\begin{array}{ccccc} G_\gamma(\alpha_{\eta,q}) & \leqslant & G_\gamma(\alpha) & \leqslant & G_\gamma(\alpha_{\eta,q+1}) \\ G_s(\alpha_{\eta,q}) & \leqslant & G_s(\alpha) & \leqslant & G_s(\alpha_{\eta,q+1}) \\ G_s(\alpha_{\eta,q+1})-\frac{G_s(\infty)}{\eta} & \leqslant & G_s(\alpha) & \leqslant & G_s(\alpha_{\eta,q})+\frac{G_s(\infty)}{\eta}. \end{array}$$

Combining these inequalities, we have

$$\begin{aligned} G_\gamma(\alpha_{\eta,q})-G_s(\alpha_{\eta,q})-\frac{G_s(\infty)}{\eta} &\leqslant G_\gamma(\alpha)-G_s(\alpha) \\ &\leqslant G_\gamma(\alpha_{\eta,q+1})-G_s(\alpha_{\eta,q+1})+\frac{G_s(\infty)}{\eta}, \end{aligned}$$

so that

$$\begin{aligned} &|G_\gamma(\alpha)-G_s(\alpha)| \\ &\quad\leqslant \max\left\{\left|G_\gamma(\alpha_{\eta,q})-G_s(\alpha_{\eta,q})\right|,\left|G_\gamma(\alpha_{\eta,q+1})-G_s(\alpha_{\eta,q+1})\right|\right\}+\frac{G_s(\infty)}{\eta} \\ &\quad\leqslant \max_{0\leqslant q'\leqslant\eta+1}\left\{\left|G_\gamma(\alpha_{\eta,q'})-G_s(\alpha_{\eta,q'})\right|\right\}+\frac{G_s(\infty)}{\eta}. \end{aligned}$$

Thus, for all $\alpha\in\mathbb{R}$,

$$|G_\gamma(\alpha)-G_s(\alpha)|^2\leqslant 2\left(\max_{0\leqslant q'\leqslant\eta+1}\left\{\left|G_\gamma(\alpha_{\eta,q'})-G_s(\alpha_{\eta,q'})\right|^2\right\}+\frac{G_s(\infty)^2}{\eta^2}\right),$$

so that

$$\mathrm{E}\left[\|G_\gamma-G_s\|_\infty^2\right]\leqslant 2\,\mathrm{E}\left[\max_{0\leqslant q\leqslant\eta+1}\left\{|G_\gamma(\alpha_{\eta,q})-G_s(\alpha_{\eta,q})|^2\right\}\right]+\frac{2G_s(\infty)^2}{\eta^2}. \tag{C.1}$$

Let $\varepsilon>0$ be given. Choose $\eta\in\mathbb{N}$ so large that $2G_s(\infty)^2\eta^{-2}<\varepsilon/2$, then use Lemma C.1 to choose Γ so that $\gamma\geqslant\Gamma$ implies

$$2\,\mathrm{E}\left[\max_{0\leqslant q\leqslant\eta+1}\left\{|G_\gamma(\alpha_{\eta,q})-G_s(\alpha_{\eta,q})|^2\right\}\right]<\frac{\varepsilon}{2}.$$

Hence, for all $\gamma\geqslant\Gamma$, the right-hand side of (C.1) is bounded by ε, which was arbitrary, so $\lim_{\gamma\to\infty}\mathrm{E}\left[\left(\|G_\gamma-G_s\|_\infty\right)^2\right]=0$. □

Proof of Theorem 3.1

By Definitions 3.1 and 3.3 and A3.0, for all $\alpha\in\mathbb{R}$,

$$F_\gamma(\alpha)=\frac{G_\gamma(\alpha)}{G_\gamma(\infty)+\mathbb{1}_{G_\gamma(\infty)=0}},\qquad F_\infty(\alpha)=\frac{G_s(\alpha)}{G_s(\infty)},$$

so

$$\begin{aligned}
\|F_\gamma - F_\infty\|_\infty &= \left\| \frac{G_\gamma}{G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0}} - \frac{G_s}{G_s(\infty)} \right\|_\infty \\
&= \left\| \frac{G_\gamma - G_s}{G_s(\infty)} + G_\gamma \frac{G_s(\infty) - (G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0})}{G_s(\infty)(G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0})} \right\|_\infty \\
&\leqslant \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{\|G_\gamma\|_\infty}{G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0}} \frac{\left|G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0} - G_s(\infty)\right|}{G_s(\infty)} \\
&\leqslant \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{\left|G_\gamma(\infty) + \mathbb{1}_{G_\gamma(\infty)=0} - G_s(\infty)\right|}{G_s(\infty)} \\
&\leqslant \frac{\|G_\gamma - G_s\|_\infty}{G_s(\infty)} + \frac{|G_s(\infty) - G_\gamma(\infty)|}{G_s(\infty)} + \frac{\mathbb{1}_{G_\gamma(\infty)=0}}{G_s(\infty)}.
\end{aligned}$$

From Lemma C.2, the first two summands converge to 0 in L_2. From A3.3e, so does the third summand. □

C.1.2 Proof of Theorem 3.2: uniform almost sure convergence of the empirical cdf

C.1.2.1 Construction of a sequence of samples

We define Y and $\mathcal{I}'_\gamma$ on the probability space $\left(\Omega \times [0,1], \mathscr{A} \otimes \mathscr{B}_{[0,1]}, \mathrm{P}' = \mathrm{P} \otimes \lambda_{[0,1]}\right)$. Define $g_\gamma(i,y) = \mathrm{P}\left(\mathcal{I}_\gamma = i | \mathcal{Y}_\gamma = y\right)$. Define $Y' : \Omega \times [0,1] \to \mathbb{R}^{\mathbb{N}}$ via

$$Y'(\omega, x) = Y(\omega).$$

Let $\mathcal{Y}'_\gamma$ be the vector of random variables $(Y'_1 \dots Y'_{N_\gamma})$ and note that $\mathcal{Y}'_\gamma(\omega, x) = \mathcal{Y}_\gamma(\omega)$. Let $S_{\gamma y} = \{i \in \mathbb{N}^{N_\gamma} : g_\gamma(i,y) \neq 0\}$ and note that for a given $y \in \mathbb{R}^{N_\gamma}$, $\sum_{i \in S_{\gamma y}} g_\gamma(i,y) = 1$. Define $h_\gamma : \mathbb{R}^{N_\gamma} \times \mathbb{N}^{N_\gamma} \to \mathbb{R}$ via

$$h_\gamma(y,i) = \sup_{\alpha \in \mathbb{R}} \left| \frac{\sum_{k \in U_\gamma} i_k \mathbb{1}_{(-\infty,\alpha]}(y_k)}{\mathbb{1}_{i=0} + \sum_{k \in U_\gamma}(i_k)} - G_s(\alpha) \right|.$$

We now impose an order on the $M_{\gamma y}$ vectors in $S_{\gamma y}$ by requiring h_γ to be non-increasing; that is, for vectors $i^{(t)}, i^{(u)} \in S_{\gamma y}$, $t < u$ if and only if $h_\gamma(y, i^{(t)}) \geqslant h_\gamma(y, i^{(u)})$. Any ties can be resolved, e.g., by randomization. For $\omega \in \Omega$ and $x \in [0,1]$, we then define $\mathcal{I}'_\gamma(\omega, 0) = i^{(1)}$ and for $x > 0$

$$\mathcal{I}'_\gamma(\omega, x) = \sum_{u=1}^{M_{\gamma y}} i^{(u)} \mathbb{1}_{\left(\sum_{t<u} g_\gamma(i^{(t)}, \mathcal{Y}_\gamma(\omega)), \sum_{t \leqslant u} g_\gamma(i^{(t)}, \mathcal{Y}_\gamma(\omega))\right]}(x).$$

Because we use uniform measure on $\mathscr{B}_{[0,1]}$, the vector $i^{(u)}$ is sampled from $S_{\gamma \mathcal{Y}_\gamma(\omega)}$ with probability $g_\gamma(i^{(u)}, \mathcal{Y}_\gamma(\omega))$. Thus, by construction we have for all γ,

$$\mathrm{P}'\left(\mathcal{I}'_\gamma = i \mid \mathcal{Y}'_\gamma = y\right) = g_\gamma(i,y) = \mathrm{P}\left(\mathcal{I}_\gamma = i \mid \mathcal{Y}_\gamma = y\right)$$

and $\mathrm{P}'\left(\mathcal{Y}'_\gamma = y\right) = \mathrm{P}\left(\mathcal{Y}_\gamma = y\right)$, so that

$$\mathrm{P}'\left(\mathcal{I}'_\gamma = i, \mathcal{Y}'_\gamma = y\right) = \mathrm{P}\left(\mathcal{I}_\gamma = i, \mathcal{Y}_\gamma = y\right).$$

This yields the following property:

Property C.1. *For all γ,*

$$h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) = \sup_{\alpha\in\mathbb{R}} \left|F'_\gamma(\alpha) - F_\infty(\alpha)\right| = \|F'_\gamma - F_\infty\|_\infty$$

has the same law as $\|F_\gamma - F_\infty\|_\infty$, where F'_γ is defined in (3.7).

Define $G'_\gamma : \mathbb{R} \to \mathbb{R}^+$ via

$$G'_\gamma(\alpha) = \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y'_k) I'_{\gamma k}}{N_\gamma},$$

noting that $F'_\gamma = G'_\gamma \left(G'_\gamma(\infty) + \mathbb{1}_{G'_\gamma(\infty)=0}\right)^{-1}$. We then have the following lemma.

Lemma C.3. *Under A3.0 and A3.2, for all $\alpha \in \mathbb{R}$,*

$$\lim_{\gamma\to\infty} \int_{[0,1]} \left(G'_\gamma(\alpha)(\omega, x) - G_s(\alpha)\right)^2 d\lambda(x) = 0 \quad \textit{P-a.s.}(\omega).$$

Proof. Let

$$\Omega_{GC} = \left\{ \omega \in \Omega : \lim_{\gamma\to\infty} \sup_{\alpha\in\mathbb{R}} \left| N_\gamma^{-1} \sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y_k)(\omega) - \int \mathbb{1}_{(-\infty,\alpha]} f \, d\lambda \right| = 0 \right\}.$$

From the Glivenko-Cantelli theorem, $\mathrm{P}(\Omega_{GC}) = 1$. We will show that for all $\omega \in \Omega_{GC}$,

$$\int_{[0,1]} \left(G'_\gamma(\alpha)(\omega, x) - G_s(\alpha)\right)^2 d\lambda(x) = o_\gamma(1).$$

Let $\omega \in \Omega_{GC}$. We then have

$$\begin{aligned}
&\sqrt{\int_{[0,1]} \left(G'_\gamma(\alpha)(\omega, x) - G_s(\alpha)\right)^2 d\lambda(x)} \\
&\leqslant \sqrt{\int_{[0,1]} \left(G'_\gamma(\alpha)(\omega, x) - \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y_k(\omega)) \int_{[0,1]} I'_{\gamma k}(\omega, u) d\lambda(u)}{N_\gamma} \right)^2 d\lambda(x)} \\
&\quad + \left| \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y_k(\omega)) \int_{[0,1]} I'_{\gamma k}(\omega, u) d\lambda(u)}{N_\gamma} \right. \\
&\qquad \left. - \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y_k(\omega)) m_\gamma(Y_k(\omega))}{N_\gamma} \right| \\
&\quad + \left| \frac{\sum_{k\in U_\gamma} \mathbb{1}_{(-\infty,\alpha]}(Y_k(\omega)) m_\gamma(Y_k(\omega))}{N_\gamma} - \int \mathbb{1}_{(-\infty,\alpha]} m_\gamma f d\lambda \right| \\
&\quad + \left| \int \mathbb{1}_{(-\infty,\alpha]} m_\gamma f d\lambda - \int \mathbb{1}_{(-\infty,\alpha]} m f d\lambda \right|.
\end{aligned}$$

The first term is the square root of

$$\mathrm{Var}\left[G'_\gamma(\alpha) \mid \mathcal{Y}'_\gamma = (Y_1(\omega), \ldots, Y_{N_\gamma}(\omega))\right] = N_\gamma^{-2} o_\gamma(N_\gamma^2) = o_\gamma(1)$$

by A3.2a. The second term is

$$\left| \sum_{k \in U_\gamma} \frac{\mathbb{1}_{(-\infty,a]}(Y_k(\omega))}{N_\gamma} \left(\mathrm{E}\left[I'_{\gamma k} \mid \mathcal{Y}'_\gamma = \left(Y_1(\omega) \dots Y_{N_\gamma}(\omega) \right) \right] - m_\gamma \left(Y_k(\omega) \right) \right) \right| = o_\gamma(1)$$

by A3.2b. The third term is $o_\gamma(1)$ because the convergence of the empirical measure given by A3.2 implies the convergence of the integral for all bounded random variables. Finally, the fourth term is $o_\gamma(1)$ by A3.0 and the dominated convergence theorem. □

The following lemma has its own interest, yielding design-based uniform L_2 convergence of the empirical cdf.

Lemma C.4. *Under A3.0 and A3.2,*

$$\int \left(h_\gamma(\mathcal{Y}'_\gamma(\omega, x), \mathcal{I}'_\gamma(\omega, x)) \right)^2 d\lambda(x) = o_\gamma(1) \quad \textit{P-a.s.}(\omega).$$

Proof. Starting from Lemma C.3 and adapting the proof of Lemma C.2, we have that: $A3.2 \Rightarrow \int \left(\| G_\gamma(\mathcal{Y}'_\gamma(\omega, x), \mathcal{I}'_\gamma(\omega, x)) - G_s \|_\infty \right)^2 d\lambda(x) = o_\gamma(1)$ P-a.s.(ω). We then adapt the end of the proof of Theorem 3.1 and get the result. □

Definition C.1. *For $\omega \in \Omega$, $\gamma \in \mathbb{N}$ and all $\varepsilon > 0$, $a_{\varepsilon,\gamma,\omega} \in [0,1]$ is defined as*

$$a_{\varepsilon,\gamma,\omega} = \int_{[0,1]} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega,x) \geqslant \varepsilon\}} d\lambda(x) \quad = \quad \lambda_{[0,1]}(\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, \cdot) \geqslant \varepsilon\}).$$

Property C.2. *For all $\varepsilon > 0$,*

$$\limsup_{\gamma \to \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega,x) > \varepsilon\}} = \mathbb{1}_{\{0\}} \quad \textit{P-a.s.}(\omega).$$

Proof. First note that $\forall x \in [0,1]$, $\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega,x) > \varepsilon\}} = \mathbb{1}_{]0, a_{\varepsilon,\gamma,\omega}]}(x)$, because by construction of $\mathcal{I}'_\gamma, \mathcal{Y}'_\gamma$, $\left\{ x \in [0,1] : h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, x) > \varepsilon \right\}$ is a subinterval of $[0,1]$ containing 0 of measure $a_{\varepsilon,\gamma,\omega}$. Further, $\forall x \in [0,1]$,

$$\limsup_{\gamma \to \infty} \mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega,x) > \varepsilon\}} = \mathbb{1}_{[0, \limsup_{\gamma \to \infty} a_{\varepsilon,\gamma,\omega}]}(x). \tag{C.2}$$

By Lemma C.4, the random variable

$$h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma)(\omega, \cdot) : ([0,1], \mathscr{B}_{[0,1]}, \lambda_{[0,1]}) \to \mathbb{R}$$

converges in $L_2(\lambda)$ to 0, P-a.s.(ω), hence it also converges in probability to 0, and so $\lim_{\gamma \to \infty} a_{\varepsilon,\gamma,\omega} = 0$. The result then follows from equation (C.2). □

Proof of Theorem 3.2

Proof. We want to show that

$$A3.0, A3.2 \Rightarrow \| F'_\gamma - F_\infty \|_\infty \xrightarrow{\text{a.s.}} 0 \text{ as } \gamma \to \infty,$$

which is equivalent to showing that

$$A3.0, A3.2 \Rightarrow \mathrm{P}'\left(\left\{ \lim_{\gamma \to \infty} h_\gamma(\mathcal{Y}'_\gamma, \mathcal{I}'_\gamma) = 0 \right\} \right) = 1.$$

Assume A3.0 and A3.2. We calculate:

$$\begin{aligned}
&\mathrm{P}'\left(\left\{\lim_{\gamma\to\infty} h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)=0\right\}\right)\\
&= \mathrm{P}'(\cap_{\varepsilon>0}\cup_\Gamma\cap_{\gamma>\Gamma}\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)<\varepsilon\})\\
&= \lim_{\varepsilon\to 0}\mathrm{P}'(\cup_\Gamma\cap_{\gamma>\Gamma}\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)<\varepsilon\})\\
&= \lim_{\varepsilon\to 0} 1-\mathrm{P}'(\cap_\Gamma\cup_{\gamma>\Gamma}\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)\geqslant\varepsilon\})\\
&= 1-\lim_{\varepsilon\to 0}\int\limsup_{\gamma\to\infty}\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x)\geqslant\varepsilon\}}\,d\mathrm{P}'(\omega,x).
\end{aligned}$$

Let $\varepsilon>0$. Applying Fubini's theorem,

$$\begin{aligned}
&\int\limsup_{\gamma\to\infty}\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x)\geqslant\varepsilon\}}\,d\mathrm{P}'(\omega,x)\\
&= \int\left(\int\limsup_{\gamma}\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x)\geqslant\varepsilon\}}\,d\lambda_{[0,1]}(x)\right)d\mathrm{P}(\omega).
\end{aligned}$$

Since we have $\limsup_{\gamma\to\infty}\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x)\geqslant\varepsilon\}}=\mathbb{1}_{\{0\}}(x)$ P-a.s.(ω), we also have for all $\varepsilon>0$ that

$$\int\limsup_{\gamma\to\infty}\mathbb{1}_{\{h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)(\omega,x)\geqslant\varepsilon\}}\,d\lambda_{[0,1]}(x)=\int_{[0,1]}\mathbb{1}_{\{0\}}(x)\,d\lambda_{[0,1]}(x)=0$$

P-a.s.(ω). Thus,

$$\mathrm{P}'\left(\left\{\lim_{\gamma\to\infty} h_\gamma(\mathcal{Y}'_\gamma,\mathcal{I}'_\gamma)=0\right\}\right)=1.$$

□

C.2 Proof of Corollaries 3.1 and 3.2

We state the following lemma which is a consequence of a theorem due to Pólya (e.g., Serfling, 1980, p. 18). The proof is omitted.

Lemma C.5. *Let* $\{u_\gamma(\cdot)\}_{\gamma\in\mathbb{N}}$ *be a sequence of increasing step functions,* $u_\gamma:\mathbb{R}\to[0,1]$, *that converges pointwise to a continuous increasing function* $u:\mathbb{R}\to[0,1]$ *with* $\lim_{y\to-\infty}u(y)=0$, $\lim_{y\to\infty}u(y)=1$ *and* $0<u(y_1)=u(y_2)<1\Rightarrow y_1=y_2$. *Define* $q_\gamma(p)=\inf\{y\in\mathbb{R}:u_\gamma(y)\geqslant p\}$, $q(p)=\inf\{y\in\mathbb{R}:u(y)\geqslant p\}$. *Then for all* K *a compact subset of* $(0,1)$, $\lim_{\gamma\to\infty}\sup_{p\in K}\{q_\gamma(p)-q(p)\}=0$.

C.2.1 Proof of Corollary 3.1

Proof. As $m_\gamma f$ and mf may have different supports, we extend the definition of ζ_∞ by

$$\forall p\in\mathbb{R},\zeta_\infty(p)=\inf\{y\in\mathbb{R}:F_\infty(y)\geqslant p\}.$$

Let K be a compact subset of $(0,1)$. Then

$$\sup_{p\in K}|\zeta_\gamma(p)-\zeta_\infty(p)|\xrightarrow[\gamma\to\infty]{P}0$$

if from all subsequences one can extract a subsequence that converges a.s. to 0. Let $\tau : \mathbb{N} \to \mathbb{N}$ be a strictly increasing function. If $\|F_\gamma - F_\infty\|_\infty \xrightarrow{L_2} 0$ then $\|F_{\tau(\gamma)} - F_\infty\|_\infty \xrightarrow{L_2} 0$ and $\|F_{\tau(\gamma)} - F_\infty\|_\infty \xrightarrow{P} 0$. Then there exists $\rho : \mathbb{N} \to \mathbb{N}$ strictly increasing such that $\|F_{\tau(\rho(\gamma))} - F_\infty\|_\infty \xrightarrow{\text{a.s.}} 0$ and by Lemma C.5, $\mathrm{P}\left(\lim_{\gamma\to\infty} \sup_{p\in K} \left|\zeta_{\tau(\rho(\gamma))}(p) - \zeta_\infty(p)\right| = 0\right) = 1$.

For the uniform L_2 convergence, let $p \in (0,1)$ and $\alpha \in \mathbb{R}$. Then $|F_\gamma(\alpha) - F_\infty(\alpha)| \leqslant \|F_\gamma - F_\infty\|_\infty$, so that

$$\begin{aligned}\{\alpha \in \mathbb{R} : F_\infty(\alpha) \geqslant p + \|F_\gamma - F_\infty\|_\infty\} &\subset \{\alpha \in \mathbb{R} : F_\gamma(\alpha) \geqslant p\} \\ &\subset \{\alpha \in \mathbb{R} : F_\infty(\alpha) \geqslant p - \|F_\gamma - F_\infty\|_\infty\},\end{aligned}$$

and

$$\begin{aligned}\inf\{\alpha \in \mathbb{R} : F_\infty(\alpha) \geqslant p + \|F_\gamma - F_\infty\|_\infty\} &\geqslant \inf\{\alpha \in \mathbb{R} : F_\gamma(\alpha) \geqslant p\} \\ &\geqslant \inf\{\alpha \in \mathbb{R} : F_\infty(\alpha) \geqslant p - \|F_\gamma - F_\infty\|_\infty\}.\end{aligned}$$

Hence $\forall p \in (0,1)$, $\zeta_\infty\left(p + \|F_\gamma - F_\infty\|_\infty\right) \geqslant \zeta_\gamma(p) \geqslant \zeta_\infty\left(p - \|F_\gamma - F_\infty\|_\infty\right)$.

Further, f has compact support by hypothesis, so there exists $b > 0$ such that the supports of $(m_\gamma f)_{\gamma\in\mathbb{N}}$ and mf are included in $[-b, b]$. So $\forall p \in (0,1)$, $\gamma \in \mathbb{N}$, $-b \leqslant \zeta_\gamma(p) \leqslant b$, $-b \leqslant \zeta_\infty(p) \leqslant b$. By combining these three inequalities, we have, $\forall p \in (0,1)$:

$$|\zeta_\infty(p) - \zeta_\gamma(p)| \leqslant \min\{b, \zeta_\infty(p + \|F_\gamma - F_\infty\|_\infty)\} - \max\{-b, \zeta_\infty(p - \|F_\gamma - F_\infty\|_\infty)\}. \tag{C.3}$$

Since $K \subset (0,1)$ is compact, there exists $a \in (0,1)$ such that $K \subset [a, 1-a]$. With the assumed continuity of F_∞, we have that ζ_∞ is uniformly continuous on any subinterval of $[0,1]$ that does not contain zero. Thus, for $\varepsilon > 0$, there exists $\eta \in (0, a/2)$ such that $p \in K$ implies $|\zeta_\infty(p+\eta) - \zeta_\infty(p-\eta)| \leqslant \varepsilon$. If $\|F_\gamma - F_\infty\|_\infty \leqslant \eta$, then $p + \|F_\gamma - F_\infty\|_\infty \leqslant p + \eta < 1 - a/2$, and $\zeta_\infty(p + \|F_\gamma - F_\infty\|_\infty) < b$, $p - \|F_\gamma - F_\infty\|_\infty \geqslant p - \eta > a/2$ and $\zeta_\infty(p - \|F_\gamma - F_\infty\|_\infty) > -b$, so equation (C.3) is bounded by ε. If $\|F_\gamma - F_\infty\|_\infty > \eta$, then (C.3) is bounded by $(2b)\mathbb{1}_{\{\|F_\gamma - F_\infty\|_\infty > \eta\}}$. Thus

$$\mathrm{E}\left[\left(\sup_{p\in K} |\zeta_\gamma(p) - \zeta_\infty(p)|\right)^2\right] \leqslant \varepsilon^2 + 4b^2\,\mathrm{P}\left(\|F_\gamma - F_\infty\|_\infty > \eta\right).$$

Since ε was arbitrary and $\mathrm{P}\left(\|F_\gamma - F_\infty\|_\infty > \eta\right) \to 0$ as $\gamma \to \infty$, the result follows.

□

C.2.2 Proof of Corollary 3.2

Proof. If $\|F'_\gamma - F_\infty\|_\infty \xrightarrow{\text{a.s.}} 0$, then for all K a compact subset of $(0,1)$, and all $(\omega, x) \in \left\{(\omega, x) : \|F'_\gamma - F_\infty\|_\infty \to 0\right\}$, we apply Lemma C.5 with $u_\gamma = F'_\gamma(\omega, x)$, $u = F_\infty$, and obtain that $\mathrm{P}'\left(\lim_{\gamma\to\infty} \sup_{p\in K} |\zeta'_\gamma(p) - \zeta'_s(p)| = 0\right) = 1$. □

C.3 Proofs for specific designs

C.3.1 Proof of A1 in the case of sampling with replacement

We consider the particular case of sampling with replacement that consists of a number of independent draws (with replacement) of units where the probability of selecting the kth individual at each draw is given by the kth coordinate of the random variable $\mathcal{Z}_\gamma : \Omega \to \{x \in \mathbb{R}^{+N_\gamma} | \sum x_k = 1\}$.

The distribution of $\mathcal{I}_\gamma$ conditional to $\mathcal{Z}_\gamma$ is $P^{\mathcal{I}_\gamma|\mathcal{Z}_\gamma=z} = h(z).\mu_{\mathbb{N}^{N_\gamma}}$ where $h(z) : i \mapsto \mathbb{1}_{\sum i_k = n_\gamma} \binom{n_\gamma}{i} z^i$. We can calculate:

$$\begin{aligned}
m_\gamma(y) &= n_\gamma \,\mathrm{E}\left[Z_{\gamma 1}|Y_{\cdot 1} = y\right] \\
m'_\gamma(y_2, y_1) &= n_\gamma \,\mathrm{E}\left[Z_{\gamma 1}|Y_{\cdot 1} = y_1, Y_{\cdot 2} = y_2\right] \\
v_\gamma(y) &= n_\gamma^2 \,\mathrm{Var}\left[Z_{\gamma 1}|Y_{\cdot 1} = y\right] + n_\gamma \,\mathrm{E}\left[Z_{\gamma 1}(1 - Z_{\gamma 1})|Y_{\cdot 1} = y\right] \\
c_\gamma(y_1, y_2) &= \mathrm{Cov}\left[\mathrm{E}\left[I_{\gamma 1}|Z_{\gamma 1}, Z_{\gamma 2}\right], \mathrm{E}\left[I_{\gamma 2}|Z_{\gamma 1}, Z_{\gamma 2}\right]|Y_{\cdot 1} = y_1, Y_{\cdot 2} = y_2\right] \\
&\quad + \mathrm{E}\left[\mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2}|Z_{\gamma 1}, Z_{\gamma 2}\right]|Y_{\cdot 1} = y_1, Y_{\cdot 2} = y_2\right] \\
&= n_\gamma^2 \,\mathrm{Cov}\left[Z_{\gamma 1}, Z_{\gamma 2}|Y_{\cdot 1} = y_1, Y_{\cdot 2} = y_2\right] \\
&\quad + n_\gamma \,\mathrm{E}\left[-n Z_{\gamma 1} Z_{\gamma 2}|Y_{\cdot 1} = y_1, Y_{\cdot 2} = y_2\right].
\end{aligned}$$

If $Z_{\gamma k} = \frac{1}{N_\gamma}$ and $\frac{n_\gamma}{N_\gamma} \to m \in [0, 1]$ then A3.1 is verified (it corresponds to the case Y and I independent).

We can consider another case: sampling with replacement with probability proportional to size: we assume that $\forall k \in \mathbb{N}$, $Y_{\cdot k} > 0$ and that $\mathcal{Z}_\gamma = \frac{Y_\gamma}{\sum_{k=1}^{N_\gamma} Y_{\cdot k}}$. With stronger conditions on the sample scheme and on Y, we can show that A3.1 holds:

Theorem C.1. *We suppose that:*

$$\begin{cases}
\forall k Y_{\cdot k} > 0 \\
\mathcal{Z}_\gamma = \frac{Y_\gamma}{\sum_{k=1}^{N_\gamma} Y_{\cdot k}} \\
\exists \tau \text{ s.t. } \frac{n_\gamma}{N_\gamma} = \tau + o_\gamma(1) \\
\mathrm{Var}\left[Y_{\cdot 1}\right] < +\infty \\
\mathrm{E}\left[Y^6\right] < +\infty.
\end{cases}$$

Then $m : y \mapsto \tau \frac{y}{\mathrm{E}[Y_{\cdot 1}]}$ *and A3.0, A3.1, A3.2 hold*

Proof. Let $U_\gamma = \frac{\sum_{k=2}^{N_\gamma} Y_{\cdot k}}{N_\gamma}$, $V_\gamma = \frac{\sum_{k=3}^{N_\gamma} Y_{\cdot k}}{N_\gamma}$. We will show that $\forall y,\ m_\gamma(y) = m(y) + o_\gamma(1)$:

$$\begin{aligned}
m_\gamma(y) &= n_\gamma \,\mathrm{E}\left[\frac{y}{y + \sum_{k=2}^{N_\gamma} Y_{\cdot k}}\right] \\
&= y \frac{n_\gamma}{N_\gamma} \,\mathrm{E}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right].
\end{aligned}$$

We want to show that: $\lim \mathrm{E}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right] = \mathrm{E}\left[Y_{\cdot 1}\right]^{-1}$. By the strong law of large numbers, we get:

$$U_\gamma \overset{L^1, L^2}{\to} \mathrm{E}\left[Y_{\cdot 1}\right].$$

Then

$$\begin{aligned}
E\frac{1}{\frac{y}{N_\gamma} + U_\gamma} - \mathrm{E}\left[Y_{\cdot 1}\right]^{-1} &= \mathrm{E}\left[Y_{\cdot 1}\right]^{-1} \mathrm{E}\left[\frac{1}{1 + \frac{y}{N_\gamma}\mathrm{E}\left[Y_{\cdot 1}\right]^{-1} + (U_\gamma - \mathrm{E}\left[Y_{\cdot 1}\right])\mathrm{E}\left[Y_{\cdot 1}\right]^{-1}} - 1\right] \\
&\leqslant \mathrm{E}\left[Y_{\cdot 1}\right]^{-1} \mathrm{E}\left[\left(\frac{y}{N_\gamma}\mathrm{E}\left[Y_{\cdot 1}\right]^{-1} + (U_\gamma - \mathrm{E}\left[Y_{\cdot 1}\right])\mathrm{E}\left[Y_{\cdot 1}\right]^{-1}\right)^2\right] \\
&= o_\gamma(1),
\end{aligned}$$

so:

$$m_\gamma(y) = \tau \frac{y}{\mathrm{E}\left[Y_{.1}\right]} + o_\gamma(1).$$

We next show that $\exists M : \mathbb{R} \to \mathbb{R}$ s.t. $M(Y) \in L_1$ and $\forall \gamma \; m_\gamma < M$. For example, $M : y \mapsto \left(\max\left\{ \frac{n_\gamma}{N_\gamma} \middle| \gamma \in \mathbb{N} \right\}\right) y$ satisfies this equation. Note that

$$\begin{aligned} m'_\gamma(y_2, y_1) &= y_1 \frac{n_\gamma}{N_\gamma} \mathrm{E}\left[\frac{1}{\frac{y_1+y_2}{N_\gamma} + V_\gamma} \right] \\ & m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) - m_\gamma(y_1) m_\gamma(y_2) \\ &= \left(\frac{n_\gamma}{N_\gamma}\right)^2 y_1 y_2 \left(\mathrm{E}\left[\frac{1}{\frac{y_1}{N_\gamma} + U_\gamma} \right] \mathrm{E}\left[\frac{1}{\frac{y_2}{N_\gamma} + U_\gamma} \right] \right. \\ & \left. - \mathrm{E}\left[\frac{1}{\frac{y_1+y_2}{N_\gamma} + V_\gamma} \right]^2 \right). \end{aligned}$$

Then, by using $\forall x > 0, \; (1-x) < (1+x)^{-1} < 1 - x + x^2$ we have:

$$\begin{aligned} \mathrm{E}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma} \right] &\leqslant \mathrm{E}\left[U_\gamma^{-1} \left(1 - \frac{y}{N_\gamma} U_\gamma^{-1} + \frac{y^2}{N_\gamma^2} U_\gamma^{-2} \right) \right] \\ \mathrm{E}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma} \right] &\geqslant \mathrm{E}\left[U_\gamma^{-1} \left(1 - \frac{y}{N_\gamma} U_\gamma^{-1} \right) \right], \end{aligned}$$

from which it follows that

$$\begin{aligned} & m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) - m_\gamma(y_1) m_\gamma(y_2) \\ \leqslant \; & \frac{y_1 y_2}{N_\gamma^2} \left(\left(\mathrm{E}\left[U_\gamma^{-1}\right]^2 - \left[V_\gamma^{-1}\right]^2 \right) + \left(\frac{y_1+y_2}{N_\gamma} E U_\gamma^{-2} \, \mathrm{E}\left[U_\gamma^{-1}\right] \right) \right. \\ & - \left(\frac{y_1+y_2}{N_\gamma} \mathrm{E}\left[V_\gamma^{-2}\right] \mathrm{E}\left[V_\gamma^{-1}\right] \right) - \left(\left(\frac{y_1+y_2}{N_\gamma} \right)^2 \mathrm{E}\left[V_\gamma^{-2}\right]^2 \right) \\ & + \left(\frac{y_1^2+y_2^2}{N_\gamma^2} \mathrm{E}\left[U_\gamma^{-3}\right] \mathrm{E}\left[U_\gamma^{-1}\right] \right) - \left(\frac{y_1 y_2}{N_\gamma^2} \mathrm{E}\left[U_\gamma^{-2}\right]^2 \right) \\ & \left. - \left(\frac{y_1 y_2^2 + y_2 y_1^2}{N_\gamma^3} \mathrm{E}\left[U_\gamma^{-2}\right] \mathrm{E}\left[U_\gamma^{-3}\right] \right) - \left(\frac{y_1^2 y_2^2}{N_\gamma^4} \mathrm{E}\left[U_\gamma^{-3}\right]^2 \right) \right) \\ \leqslant \; & \frac{y_1 y_2}{N_\gamma^2} \left(o_\gamma(1) + \frac{y_1+y_2}{N_\gamma} o_\gamma(1) - \frac{(y_1+y_2)^2}{N_\gamma^2} O_\gamma(1) \right. \\ & \left. + \frac{y_1^2+y_2^2}{N_\gamma^2} O_\gamma(1) - \frac{y_1 y_2}{N_\gamma^2} O_\gamma(1) - \frac{y_1 y_2^2 + y_2 y_1^2}{N_\gamma^3} O_\gamma(1) - \frac{y_1^2 y_2^2}{N_\gamma^4} O_\gamma(1) \right). \end{aligned}$$

We then conclude that

$$\int \left(m'_\gamma(y_1, y_2) m'_\gamma(y_2, y_1) - m_\gamma(y_1) m_\gamma(y_2) \right) f(y_1) f(y_2) dy_1 dy_2 = o_\gamma\left(1/N_\gamma^2\right).$$

We next show that $\int v_\gamma \, f \, d\lambda = o_\gamma(N_\gamma)$:

$$\begin{aligned}
v_\gamma(y) &= n_\gamma^2 \operatorname{Var}\left[\frac{y}{y + N_\gamma U_\gamma}\right] + n_\gamma \operatorname{E}\left[\frac{y N_\gamma U_\gamma}{(y + N_\gamma U_\gamma)^2}\right] \\
&= \left(\frac{n_\gamma}{N_\gamma} y\right)^2 \operatorname{Var}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right] + \frac{n_\gamma}{N_\gamma}\frac{y}{N_\gamma} \operatorname{E}\left[\frac{U_\gamma}{\left(\frac{y}{N_\gamma} + U_\gamma\right)^2}\right] \\
\left|n_\gamma \operatorname{E}\left[\frac{\frac{y}{N_\gamma} U_\gamma}{\left(\frac{y}{N_\gamma} + U_\gamma\right)^2}\right]\right| &\leqslant \left(\frac{n_\gamma}{N_\gamma} y\right) \operatorname{E}\left[\frac{N_\gamma}{U_\gamma}\right] \\
&\leqslant \left(\frac{n_\gamma}{N_\gamma} y\right)\left(\frac{1}{\operatorname{E}[Y_{\cdot 1}]} + o_\gamma(1)\right) \\
\left|\int n_\gamma \operatorname{E}\left[\frac{y N_\gamma U_\gamma}{(y + N_\gamma U_\gamma)^2}\right] f(y) dy\right| &\leqslant \frac{n_\gamma}{N_\gamma} + o_\gamma(1) \\
&= o_\gamma\left(N_\gamma\right).
\end{aligned}$$

Now we use the inequalities:

$$\forall x, y > 0, \ (y + x)^{-1} < x^{-1}, \ (1 + x)^{-1} > 1 - x.$$

$$\begin{aligned}
\operatorname{Var}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right] &= \operatorname{E}\left[\left(\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right)^2\right] - \operatorname{E}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right]^2 \\
&\leqslant \operatorname{E}\left[\frac{1}{U_\gamma^2}\right] - \operatorname{E}\left[U_\gamma^{-1}\left(1 - \frac{y}{N_\gamma} U_\gamma^{-1}\right)\right]^2 \\
&= \left(\operatorname{E}[Y_{\cdot 1}]^{-2} + o_\gamma(1)\right) - \\
&\qquad - \left(\operatorname{E}[Y_{\cdot 1}]^{-1} - \frac{y}{N_\gamma} \operatorname{E}[Y_{\cdot 1}]^{-2} + o_\gamma(1)\right)^2 \\
&= 2\frac{y}{N_\gamma} \operatorname{E}[Y_{\cdot 1}]^{-3} - \left(\frac{y}{N_\gamma}\right)^2 \operatorname{E}[Y_{\cdot 1}]^{-4} + o_\gamma(1),
\end{aligned}$$

where the $o_\gamma()$ do not depend on y.

Then

$$\begin{aligned}
\int \left(\frac{n_\gamma}{N_\gamma} y\right)^2 \operatorname{Var}\left[\frac{1}{\frac{y}{N_\gamma} + U_\gamma}\right] f(y) dy &\leqslant \left(\frac{n_\gamma}{N_\gamma}\right)^2 \left(\frac{2 \operatorname{E}\left[Y_{\cdot 1}^3\right] \operatorname{E}[Y_{\cdot 1}]^{-3}}{N_\gamma} - \frac{\operatorname{E}\left[Y_{\cdot 1}^4\right] \operatorname{E}[Y_{\cdot 1}]^{-4}}{N_\gamma^2} + o_\gamma(1)\right) \\
&= o_\gamma(1),
\end{aligned}$$

so

$$\int v_\gamma \, f \, d\lambda = o_\gamma(N_\gamma).$$

Next,

$$\begin{aligned} c_\gamma(y_1,y_2) &= -n_\gamma \,\mathrm{E}\left[\frac{y_1y_2}{(y_1+y_2+N_\gamma V_\gamma)^2}\right] + n_\gamma^2 \,\mathrm{Cov}\left[\frac{y_1}{y_1+y_2+N_\gamma V_\gamma}, \frac{y_2}{y_1+y_2+N_\gamma V_\gamma}\right] \\ &= \left(\frac{n_\gamma}{N_\gamma}\right)^2 y_1y_2 \left(-\frac{1}{N_\gamma}\,\mathrm{E}\left[\frac{1}{\left(\frac{y_1+y_2}{N_\gamma}+V_\gamma\right)^2}\right] + n_\gamma \,\mathrm{Var}\left[\frac{1}{\frac{y_1+y_2}{N_\gamma}+V_\gamma}\right]\right), \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\left[\frac{1}{\left(\frac{y_1+y_2}{N_\gamma}+V_\gamma\right)^2}\right] &\leqslant \mathrm{E}\left[V_\gamma^{-2}\right], \\ \mathrm{Var}\left[\frac{1}{\frac{y_1+y_2}{N_\gamma}+V_\gamma}\right] &\leqslant 2\frac{y_1+y_2}{N_\gamma}\,\mathrm{E}\left[Y_{.1}\right]^{-3} - \left(\frac{y_1+y_2}{N_\gamma}\right)^2 \mathrm{E}\left[Y_{.1}\right]^{-4} + o_\gamma(1), \end{aligned}$$

where the $o_\gamma()$ do not depend on y_1, y_2.

It follows that

$$\begin{aligned} \int c_\gamma(y_1,y_2)f(y_1)f(y_2)dy_1dy_2 &\leqslant O_\gamma\left(\frac{1}{n_\gamma}\right) \\ &\quad +2\left(\frac{n_\gamma}{N_\gamma}\right)^2\left(\frac{\mathrm{E}\left[Y_{.1}Y_{.2}(Y_{.1}+Y_{.2})\right]\mathrm{E}\left[Y_{.1}\right]^{-3}}{N_\gamma}\right. \\ &\quad \left.-\left(\frac{\mathrm{E}\left[Y_{.1}Y_{.2}(Y_{.1}+Y_{.2})^2\right]\mathrm{E}\left[Y_{.1}\right]^{-4}}{N_\gamma}\right)^2\right) + o_\gamma(1) \\ &= o_\gamma(1). \end{aligned}$$

□

Now we prove that A3.2 holds under the same conditions: Let $\mathrm{Var}\left[\mathcal{I}_\gamma|Y=y\right]_{kl}$ denote the (k,l)th element of the conditional variance-covariance matrix of $\mathcal{I}_\gamma$.

Proof.

$$\begin{aligned} &\text{For } k\neq l,\ \mathrm{Var}\left[\mathcal{I}_\gamma|Y=y\right]_{kl} \\ &= -n_\gamma\frac{y_ky_l}{\left(\sum_{h=1}^{N_\gamma}y_h\right)^2} \\ &< 0 \\ &\mathrm{Var}\left[\mathcal{I}_\gamma|Y=y\right]_{kk} \\ &= n_\gamma\frac{y_k\left(\sum_{h\in[\![1,N_\gamma]\!]\setminus\{k\}}y_l\right)}{\left(\sum_{h=1}^{N_\gamma}y_h\right)^2} \end{aligned}$$

For $\alpha \in \mathbb{R}$

$$\begin{aligned}
&\operatorname{Var}\left[\sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_k) I_{\gamma k} | Y_\gamma = (y_1 \dots y_{N_\gamma})\right] \\
&\leqslant \sum_{k=1}^{N_\gamma} b_\alpha^2(y_k) n_\gamma \frac{y_k \left(\sum_{h \in [\![1,N_\gamma]\!] \setminus \{k\}} y_l\right)}{\left(\sum_{h=1}^{N_\gamma} y_h\right)^2} \\
&\leqslant n_\gamma \frac{\sum_{k=1}^{N_\gamma} y_k \left(\sum_{h=1}^{N_\gamma} y_h\right) - \sum_{k=1}^{N_\gamma} y_k^2}{\left(\sum_{h=1}^{N_\gamma} y_h\right)^2} \\
&\leqslant n_\gamma \left(1 - \frac{\sum_{k=1}^{N_\gamma} y_k^2}{\left(\sum_{h=1}^{N_\gamma} y_h\right)^2}\right) \\
&\leqslant n_\gamma \\
&= o(N_\gamma^2).
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) \left(\mathrm{E}\left[I_{\gamma k} | Y_\gamma = (y_1 \dots y_{N_\gamma})\right] - m_\gamma(y_k)\right) \\
&= \sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) n_\gamma y_k \left(\frac{1}{\sum_{h=1}^{N_\gamma} y_h} - \frac{1}{N_\gamma} \mathrm{E}\left[\frac{1}{\frac{y_k}{N_\gamma} + U_\gamma}\right]\right) \\
&\in \left[\sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) n_\gamma y_k \left(\frac{1}{\sum_{h=1}^{N_\gamma} y_h} - \frac{1}{N_\gamma \mathrm{E}\left[Y_{\cdot 1}\right]}\right.\right. \\
&\qquad \left.\left. \mp \frac{1}{N_\gamma} \mathrm{E}\left[Y_{\cdot 1}\right]^{-3} \mathrm{E}\left[\left(\frac{y_k}{N_\gamma} + (U_\gamma - \mathrm{E}\left[Y_{\cdot 1}\right])\right)^2\right]\right)\right] \\
&\subset \sum_{k=1}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(y_k) n_\gamma y_k \left(\frac{1}{\sum_{h=1}^{N_\gamma} y_h} - \frac{1}{N_\gamma \mathrm{E}\left[Y_{\cdot 1}\right]}\right) \\
&\qquad + \left[\mp n_\gamma \alpha \mathrm{E}\left[Y_{\cdot 1}\right]^{-3} \mathrm{E}\left[\left(\frac{\alpha}{N_\gamma} + (U_\gamma - \mathrm{E}\left[Y_{\cdot 1}\right])\right)^2\right]\right] \\
&\subset n_\gamma O_\gamma(1) + \left[\mp n_\gamma O_\gamma(1)\right].
\end{aligned}$$

□

C.3.2 Proof for stratified simple random sampling without replacement, with non random number of strata stratum sizes, and stratum sample sizes

Let X_γ be a discrete variable with value in $[\![1, H_\gamma]\!]^{N_\gamma}$. For each γ, let $N_{\gamma 1}, \dots N_{\gamma H_\gamma}$ be a finite sequence of integers such that $\forall h, \#\{k | X_{\gamma k} = h\} = N_{\gamma h}$. Let $n_\gamma = \left(n_{\gamma 1}, \dots n_{\gamma H_\gamma}\right)$ be a random vector of integers such that $n_{\gamma h} \leqslant N_{\gamma h}$. Given (X_γ, n_γ), the sample is selected via a stratified sample with SRS of $n_{\gamma h}$ from

$N_{\gamma h}$ elements within the stratum h, and independence between strata, e.g.

$$\Pi_\gamma(i) = \begin{vmatrix} \frac{1}{\prod_{h=1}^{H_\gamma} \binom{n_\gamma}{N_\gamma}} & \text{if } \forall h,\ \sum_{k|X_{\gamma k}=h} i_k = n_{\gamma h} \text{ and } i \in \{0,1\}^{N_\gamma} \\ 0 & \text{otherwise.} \end{vmatrix}$$

Introduce the random variable (the order statistic) a.s. defined η_γ the permutation of $[\![1, N_\gamma]\!]$ such that $Y_{\eta_\gamma(1)} < \ldots < Y_{\eta_\gamma(N_\gamma)}$ and consider the case where X_γ is ordered like Y_γ:

$$X_{\gamma k} = \sum_{h=1}^{N_\gamma} \mathbb{1}_{\sum_{g<h} N_{\gamma g} < \eta_\gamma(k) \leqslant \sum_{g \leqslant h} N_{\gamma g}}$$

Theorem C.2. *If*

$$\tau : \alpha = \lim_\gamma \sum_{h=0}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathbb{1}_{\left]\frac{\sum_{g<h} N_{\gamma g}}{N_\gamma}, \frac{\sum_{g \leqslant h} N_{\gamma g}}{N_\gamma}\right[}(\alpha)$$

exists except for a finite number of points and is a piecewise-continuous non null function, and the limit is uniform in α on the set of continuity points of τ, then A3.3 and A3.2 hold.

Proof. • We first show that A3.3a holds

As H_γ, N_γ and n_γ are not random,

$$m_\gamma(y) = \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(X_{\gamma 1} = h | Y_{\gamma 1} = y\right).$$

Introduce the empirical cumulative distribution function given by the $N-1$ (resp $N-2$) $Y_{.k}$: $D_{\gamma 1} : \alpha \mapsto \frac{\sum_{k=2}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_{.k})}{N_\gamma}$ (resp. $D_{\gamma 2} : \alpha \mapsto \frac{\sum_{k=3}^{N_\gamma} \mathbb{1}_{]-\infty,\alpha]}(Y_{.k})}{N_\gamma}$). From the classic Glivenko-Cantelli theorem, $D_{\gamma 1}$ and $D_{\gamma,2}$ converge to F uniformly almost surely.

Then the pair $(\alpha_1, \alpha_2) \mapsto D_{\gamma 1}(\alpha_1), D_{\gamma 2}(\alpha_2)$ do also uniformly converge to $(\alpha_1, \alpha_2) \mapsto F(\alpha_1), F(\alpha_2)$, and

$$\begin{aligned}
&P\left(X_{\gamma 1} = h | Y_{\gamma 1} = y\right) \\
&= P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} \leqslant D_{\gamma 1}(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma}\right) \\
&P\left(X_{\gamma 2} = h | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right) \\
&= P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} \leqslant D_{\gamma 2}(y_2) + \frac{\mathbb{1}_{y_2 > y_1}}{N_\gamma} < \sum_{g=0}^{h} \frac{N_{\gamma g}}{N_\gamma}\right) \\
&P\left(X_{\gamma 1} = h_1, X_{\gamma 1} = h_2 | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right) \\
&= P\left(\left\{\frac{\sum_{g=0}^{h_2-1} N_{\gamma g}}{N_\gamma} \leqslant D_{\gamma 2}(y_2) + \frac{\mathbb{1}_{y_2 > y_1}}{N_\gamma} < \sum_{g=0}^{h_2} \frac{N_{\gamma g}}{N_\gamma}\right\}\right. \\
&\qquad \left. \cap \left\{\frac{\sum_{g=0}^{h_1-1} N_{\gamma g}}{N_\gamma} \leqslant D_{\gamma 2}(y_1) + \frac{\mathbb{1}_{y_1 > y_2}}{N_\gamma} < \sum_{g=0}^{h_1} \frac{N_{\gamma g}}{N_\gamma}\right\}\right).
\end{aligned}$$

We show that for y s.t. τ is continuous in $F(y)$, $\lim_\gamma m_\gamma(y) = \tau(F(y))$; i.e. that $\forall \varepsilon \in \mathbb{R}^{*+}$, $\exists \Gamma \in \mathbb{N}$ s.t. $\gamma > \Gamma \Rightarrow |m_\gamma(y) - \tau(F(y))| < \varepsilon$: Let y s.t. τ is continuous in $F(y)$, $\varepsilon \in \mathbb{R}^{*+}$, and let $\varepsilon_1, \varepsilon_2, \varepsilon_3$ such that $\varepsilon_3 \tau(F(y)) + \varepsilon_1 + \varepsilon_2 + \varepsilon_3(\varepsilon_1 + \varepsilon_2) < \varepsilon$. Then $\exists \nu > 0$ such that $\|\tau_{|[F(y)-\nu, F(y)+\nu]} - \tau(F(y))\|_\infty < \varepsilon_1$ and $\exists \Gamma_1$ such that $\forall \gamma > \Gamma_1$, $\forall \alpha \in [F(y) - \nu, F(y) + \nu]$,

$$\sum_{h=0}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \mathbb{1}_{\left]\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_{\gamma g}}, \frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_{\gamma g}}\right[}(\alpha) - \tau(\alpha) < \varepsilon_2$$

In addition, there exists Γ_2 such that $\forall \gamma > \Gamma_2$ $P(|F_(y) - D_{\gamma 1}(y)| > \nu) < \varepsilon_3$, so for $\gamma > \max\{\Gamma_1, \Gamma_2\}$,

$$\begin{aligned} m_\gamma(y) &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(X_{\gamma 1} = h | Y_{\gamma 1} = y\right) \\ &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right). \end{aligned}$$

Let $\Delta_\gamma = \left\{h \in [\![1, H_\gamma]\!] \,|\, \sum_{g=0}^{h-1} \frac{N_{\gamma g}}{N_\gamma} < F(y) + \nu, \sum_{g=0}^{h} \frac{N_{\gamma g}}{N_\gamma} > F(y) - \nu\right\}$. Then for $h \notin \Delta_\gamma$:

$$\begin{aligned} 0 &\leqslant \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\ &\leqslant P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right), \end{aligned}$$

so:

$$\begin{aligned} &\sum_{h \in \Delta_\gamma}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\ &\leqslant m_\gamma(y) \\ &\leqslant \sum_{h \notin \Delta_\gamma}^{H_\gamma} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\ &\quad \sum_{h \in \Delta_\gamma}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\ &\leqslant \varepsilon_3 + \sum_{h \in \Delta_\gamma}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \end{aligned}$$

For $h \in \Delta_\gamma$, $\tau(F(y)) - \varepsilon_1 - \varepsilon_2 \leqslant \frac{n_{\gamma h}}{N_{\gamma h}} \leqslant \tau(F(y)) + \varepsilon_1 + \varepsilon_1$, so:

$$\begin{aligned}
&\tau(F(y)) - \varepsilon \\
&\quad\leqslant\quad (1-\varepsilon_3)(\tau(F(y)) - \varepsilon_1 - \varepsilon_2) \\
&\quad\leqslant\quad \sum_{h\in\Delta_\gamma}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\
&\quad\leqslant\quad m_\gamma(y) \\
&\quad\leqslant\quad \varepsilon_3 + \sum_{h\in\Delta_\gamma}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(\frac{\sum_{g=0}^{h-1} N_{\gamma g}}{N_\gamma} - F(y) \leqslant D_{\gamma 1}(y) - F(y) < \frac{\sum_{g=0}^{h} N_{\gamma g}}{N_\gamma} - F(y)\right) \\
&\quad\leqslant\quad \tau(F(y)) + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\
&\quad\leqslant\quad \tau(F(y)) + \varepsilon,
\end{aligned}$$

and so A3.3a holds

- With quite the same proof, we can show that $\lim_\gamma m'_\gamma(y_1, y_2) = \tau(F(y_2))$ and $\lim_\gamma c_\gamma(y_1, y_2) = 0$ with

$$\begin{aligned}
m'_\gamma(y_1, y_2) &= \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} P\left(X_{\gamma 2} = h | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right) \\
c_\gamma(y_1, y_2) &= \mathrm{Cov}\left[\mathrm{E}\left[I_{\gamma 1} | X_{\gamma 1}, X_{\gamma 2}, Y_{\gamma 1}, Y_{\gamma 2}\right], \mathrm{E}\left[I_{\gamma 2} | X_{\gamma 1}, X_{\gamma 2}, Y_{\gamma 1}, Y_{\gamma 2}\right] | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right] \\
&\quad + \mathrm{E}\left[\mathrm{Cov}\left[I_{\gamma 1}, I_{\gamma 2} | X_{\gamma 1}, X_{\gamma 2}, Y_{\gamma 1}, Y_{\gamma 2}\right] | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right] \\
&= \sum_{h,h'=1}^{H_\gamma} \frac{n_{\gamma h}}{N_{\gamma h}} \frac{n_{\gamma h'}}{N_{\gamma h'}} \left(P\left(X_{\gamma 1} = h, X_{\gamma 2} = h' | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right)\right. \\
&\quad \left. - P\left(X_{\gamma 1} = h | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right) P\left(X_{\gamma 2} = h' | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right)\right) \\
&\quad + \sum_{h=1}^{H_\gamma} \frac{n_{\gamma h}(n_{\gamma h} - N_{\gamma h})}{N_{\gamma h}^2 (N_{\gamma h} - 1)} P\left(X_{\gamma 1} = X_{\gamma 2} = h | Y_{\gamma 1} = y_1, Y_{\gamma 2} = y_2\right),
\end{aligned}$$

so that A3.3b and A3.3c hold.

- $\forall \gamma, P(\mathcal{I}_\gamma = 0) = 0$ so A3.3d holds.

We can adapt the proof, which is based on an almost sure result, to show A3.2 □

References

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Appendix D

Proofs for chapter 5

D.1 Proof of Lemma 5.1

We begin the poof of Lemma 5.1 with the following lemma:

Lemma D.1. *Let $u : (\mathbb{R}^p, \mathscr{B}_p) \to (\mathbb{R}^p, \mathscr{B}_p)$ a measurable function, $\theta \in \Theta, \xi \in \Xi$. If u and $\mathbb{1}$ satisfy A5.1 for θ, ξ, then*

$$\frac{\sum_{k=1}^{N_\gamma} u(Y_{\gamma k}) I_{\gamma k}}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} \xrightarrow[\gamma\to\infty]{\mathrm{P}_{\theta,\xi}} \int u \rho_{\infty,\theta,\xi} f_{\theta_0} \, d\lambda_{\mathbb{R}^p}.$$

Proof. If u satisfies A5.1 for θ, ξ, then $\lim_{\gamma\to\infty} \mathrm{E}_{\theta,\xi}\left[\frac{\sum_{k=1}^{N_\gamma} u(Y_{\gamma k}) I_{\gamma k}}{N_\gamma}\right] = \mathrm{E}_{\theta,\xi}[I_{\gamma k}] \int u \rho_{\infty,\theta,\xi} f_{\theta_0} \, d\lambda_{\mathbb{R}^p}$

If u satisfies A5.1a, A5.1b, A5.1c, then $\lim_{\gamma\to\infty} \mathrm{Var}_{\theta,\xi}\left[\frac{\sum_{k=1}^{N_\gamma} u(Y_{\gamma k}) I_{\gamma k}}{N_\gamma}\right] = 0$. So if u and $\mathbb{1}_{\mathbb{R}^p}$ satisfy A5.1 then

$$\begin{aligned}
&\left(\frac{1}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} \sum_{k=1}^{N_\gamma} I_{\gamma k} u(Y_{\gamma k})\right) - \int u \rho_{\infty,\theta,\xi} f_\theta d\lambda \\
&= \left(\frac{1}{N_\gamma \int m_{\gamma,\theta,\xi} f_\theta(y) \, d\lambda_{\mathbb{R}^p}(y)} \sum_{k=1}^{N_\gamma} I_{\gamma k} u(Y_{\gamma k})\right) - \int u \rho_{\infty,\theta,\xi} f_\theta d\lambda \\
&\quad + \left(\frac{1}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} - \frac{1}{N_\gamma \int m_{\gamma,\theta,\xi} f_\theta(y) \, d\lambda_{\mathbb{R}^p}(y)}\right)\left(\sum_{k=1}^{N_\gamma} I_{\gamma k} u(Y_{\gamma k})\right) \\
&= o_{\mathrm{P}_{\theta,\xi}}(1) + \left(o_{\mathrm{P}_{\theta,\xi}}(1)\right)\left(\int \Delta(y,\xi) \rho_{\infty,\theta,\xi} f_\theta(y) \, d\lambda_{\mathbb{R}^p}(y) + o_{\mathrm{P}_{\theta,\xi}}(1)\right) \\
&= o_{\mathrm{P}_{\theta,\xi}}(1).
\end{aligned}$$

□

We now prove Lemma 5.1. Consider the Taylor series expansion

$$\frac{1}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} \sum_{k=1}^{N_\gamma} I_{\gamma k} \Delta(Y_{\gamma k}, \hat{\xi}_\gamma) \tag{D.1}$$

$$= \frac{\sum_{k=1}^{N_\gamma} I_{\gamma k} \Delta(Y_{\gamma k}, \xi_0)}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} + \left(\hat{\xi}_\gamma - \xi_0\right) \frac{1}{n_\gamma + \mathbb{1}_{\{0\}}(n_\gamma)} \sum_{k=1}^{N_\gamma} I_{\gamma k} \frac{\partial}{\partial \xi} \Delta\left(Y_{\gamma k}, \tilde{\xi}_\gamma\right) \tag{D.2}$$

with $\tilde{\xi}_\gamma \in \left[\xi_0, \hat{\xi}_\gamma\right]$. From Lemma D.1,

$$\frac{\sum_{k=1}^{N_\gamma} I_{\gamma k} \Delta\left(Y_{\gamma k}, \xi_0\right)}{n_\gamma + \mathbb{1}_{\{0\}}\left(n_\gamma\right)} = \int \Delta(y, \xi_0) \rho_{\infty,\theta_0,\xi_0}(y) f_\theta(y) d\lambda + o_{\mathrm{P}_{\theta_0,\xi_0}}(1).$$

Let $\varepsilon, d \in]0, +\infty[$, and define: $S_\gamma = \left\{\omega \in \Omega | \hat{\xi}_\gamma(\omega) \in B\right\}$ and $T_\gamma = \left\{\omega \in \Omega | \left(\left(n_\gamma + \mathbb{1}_{\{0\}}\left(n_\gamma\right)\right)^{-1} \sum_{k=1}^{N_\gamma} I_{\gamma k} R\left(Y_{\gamma k}\right)\right)(\omega) \leqslant\right.$

There exists an integer Γ such that $\gamma > \Gamma \Rightarrow \mathrm{P}_{\theta_0,\xi_0}\left(S_\gamma\right) > 1 - \frac{\varepsilon}{2}$ and $\mathrm{P}_{\theta_0,\xi_0}\left(T_\gamma\right) > 1 - \frac{\varepsilon}{2}$. We then have $\forall \gamma > \Gamma, \mathrm{P}_{\theta_0,\xi_0}\left(\left|\frac{1}{n_\gamma + \mathbb{1}_{\{0\}}\left(n_\gamma\right)} \sum_{k=1}^{N_\gamma} I_{\gamma k} \frac{\partial}{\partial \xi} \Delta\left(Y_{\gamma k}, \tilde{\xi}_\gamma\right)\right| \leqslant \int R \rho_{\infty,\theta_0,\xi_0} f_{\theta_0} d\lambda_{\mathbb{R}^p} + d\right) \geqslant \mathrm{P}_{\theta_0,\xi_0}\left(S_\gamma \cap T_\gamma\right) > 1 - \varepsilon$ so

$$\frac{1}{n_\gamma + \mathbb{1}_{\{0\}}\left(n_\gamma\right)} \sum_{k=1}^{N_\gamma} I_{\gamma k} \frac{\partial}{\partial \xi} \Delta\left(Y_{\gamma k}, \tilde{\xi}_\gamma\right) = O_{\mathrm{P}_{\theta_0 \xi_0}}(1), \tag{D.3}$$

completing the proof.

D.2 Proof of Theorem 5.1

Let $\theta \in A$ and define: $\Lambda_\theta : \mathbb{R}^p \to \mathbb{R}, (y, \xi) \mapsto \Delta(y, \theta, \xi) - \Delta(y, \theta_0, \xi) = \ln\left(\frac{\rho_{\infty,\theta,\xi}(y) f_\theta(y)}{\rho_{\infty\theta_0\xi}(y) f_{\theta_0}(y)}\right)$. By A5.2b, Λ is differentiable in ξ $\forall \xi \in B$, and by A5.2e, $\forall \xi \in B, \lambda_{\mathbb{R}^p} - a.s.(y), \left|\frac{\partial \Lambda_\theta}{\partial \xi}(y, \xi)\right| \leqslant K(y, \theta)$. Thus, by Lemma 5.1,

$$\begin{aligned} &\overline{\mathscr{L}_\gamma}\left(\theta, \hat{\xi}_\gamma, \left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}\right) - \overline{\mathscr{L}_\gamma}\left(\theta_0, \hat{\xi}_\gamma, \left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}\right) \\ &\xrightarrow[\gamma \to \infty]{\mathrm{P}_{\theta_0,\xi_0}} \int_{\theta_0,\xi_0} \ln\left(\frac{\rho_{\infty,\theta,\xi_0}(y) f_\theta(y)}{\rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y)}\right) \rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y) d\lambda_{\mathbb{R}^p}(y). \end{aligned}$$

From Jensen's inequality and A5.2g,

$$\begin{aligned} &\int_{\theta_0,\xi_0} \ln\left(\frac{\rho_{\infty,\theta,\xi_0}(y) f_\theta(y)}{\rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y)}\right) \rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y) d\lambda_{\mathbb{R}^p}(y) \\ &< \ln\left(\int_{\theta_0,\xi_0} \frac{\rho_{\infty,\theta,\xi_0}(y) f_\theta(y)}{\rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y)} \rho_{\infty,\theta_0,\xi_0}(y) f_{\theta_0}(y) d\lambda_{\mathbb{R}^p}(y)\right) = 0. \end{aligned}$$

Let $\varepsilon, d \in]0, +\infty[$ such that $]\theta_0 - \varepsilon, \theta_0 + \varepsilon[\subset A$, and let $\Gamma \in \mathbb{N}$ such that $\gamma > \Gamma \Rightarrow \mathrm{P}_{\theta_0,\xi_0}\left(\hat{\xi}_\gamma \in B\right) > 1 - d$ and $\mathrm{P}_{\theta_0,\xi_0}\left(\overline{\mathscr{L}_\gamma}\left(\theta, \hat{\xi}_\gamma, \left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}\right) < \overline{\mathscr{L}_\gamma}\left(\theta_0, \hat{\xi}_\gamma, \left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}\right)\right) > 1 - d$. Thus, for $\gamma > \Gamma$, $\mathrm{P}_{\theta_0,\xi_0}\left(\theta \mapsto \overline{\mathscr{L}_\gamma}\left(\theta, \hat{\xi}_\gamma, \left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}\right) \text{ has a local maximum } \hat{\theta}_\gamma \in]\theta_0 - \varepsilon, \theta_0 + \varepsilon[\right) > 1 - 3d$. By $A1.a$ and $A1.b$, $\frac{\partial \overline{\mathscr{L}}}{\partial \theta}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k \in \{1,\ldots,n_\gamma\}}, \hat{\theta}_\gamma, \hat{\xi}_\gamma\right) = 0$.

D.3 Proof of Theorem 5.2

By condition A5.2 and the consistency of $\hat{\theta}_\gamma$ and $\hat{\xi}_\gamma$, we can use Taylor to expand

$$\begin{aligned}
0 &= \sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\hat{\theta}_\gamma,\hat{\xi}_\gamma\right)\\
&= \sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\\
&\quad+\sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)\frac{\partial^2}{\partial\theta^2}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\\
&\quad+\sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)^2\frac{1}{2}\frac{\partial^3}{\partial\theta^3}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\tilde{\theta}_\gamma,\hat{\xi}_\gamma\right)\\
&\quad+o_{\mathrm{P}_{\theta_0,\xi_0}}(1).
\end{aligned}$$

where $\tilde{\theta}_\gamma\in[\theta,\hat{\theta}_\gamma]$, and rearrange to obtain

$$\begin{aligned}
&-\sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\\
&= \sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)\left(\frac{\partial^2\bar{\mathscr{L}}}{\partial\theta^2}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\right.\\
&\qquad\left.+\left(\hat{\theta}_\gamma-\theta_0\right)\frac{1}{2}\frac{\partial^3\bar{\mathscr{L}}}{\partial\theta^3}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\tilde{\theta}_\gamma,\hat{\xi}_\gamma\right)\right)\\
&\qquad+o_{\mathrm{P}_{\theta_0,\xi_0}}(1).
\end{aligned}$$

We examine several terms in the previous equation separately. Using conditions A5.2b, A5.2c,A5.2d, A5.2f, A5.2h and the consistency of $\hat{\theta}_\gamma$ and $\hat{\xi}_\gamma$, we establish the following identities (D.7), (D.8) and (D.9).

First, we expand $\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)$ about ξ_0, yelding

$$\begin{aligned}
&\sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\\
&= \sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)+\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)\frac{\partial^2}{\partial\theta\partial\xi}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)\\
&\quad+\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)^2\frac{\partial^3}{\partial\theta\partial\xi^2}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)+o_{\mathrm{P}_{\theta_0,\xi_0}}(1). \qquad \text{(D.4)}
\end{aligned}$$

Arguing as in Lemma 5.1, one can show that $\frac{\partial^3}{\partial\theta\partial\xi^2}\bar{\mathscr{L}}$ is bounded in probability. This establishes that

$$\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)^2\frac{\partial^3}{\partial\theta\partial\xi^2}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)=o_{\mathrm{P}_{\theta_0,\xi_0}}(1). \qquad \text{(D.5)}$$

Next, A5.1 is satisfied for $u=\frac{\partial^2}{\partial\theta\partial\xi}\Lambda$, so that

$$\left(\frac{\partial^2}{\partial\theta\partial\xi}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)+\mathscr{I}_{12}\right)=o_{\mathrm{P}_{\theta_0,\xi_0}}(1),$$

and

$$\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)\left(\frac{\partial^2}{\partial\theta\partial\xi}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)+\mathscr{I}_{12}\right)=o_{\mathrm{P}_{\theta_0,\xi_0}}(1). \qquad \text{(D.6)}$$

Next, from condition A5.2f, and Lemma 5.1.

$$\frac{\partial^2}{\partial\theta^2}\bar{\mathscr{L}}\left(\theta_0,\hat{\xi}_\gamma\right)+\mathscr{I}_{11}=o_{\mathrm{P}_{\theta_0,\xi_0}}(1). \tag{D.7}$$

Combining (D.4)-(D.7), we have

$$\begin{aligned}&\sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\hat{\xi}_\gamma\right)\\ &=\sqrt{n_\gamma}\frac{\partial}{\partial\theta}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)-\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)\mathscr{I}_{12}+o_{\mathrm{P}_{\theta_0,\xi_0}}(1).\end{aligned} \tag{D.8}$$

Further, $\frac{\partial^3}{\partial\theta^3}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\tilde{\theta}_\gamma,\hat{\xi}_\gamma\right)$ is bounded in probability, by condition A5.2f and an argument similar to that made in proving Lemma 5.1. It then follows that

$$\frac{1}{2}\left(\hat{\theta}_\gamma-\theta_0\right)\frac{\partial^3}{\partial\theta^3}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\tilde{\theta}_\gamma,\hat{\xi}_\gamma\right)=o_{\mathrm{P}_{\theta_0,\xi_0}}(1). \tag{D.9}$$

Applying the three identities above, we see that $\sqrt{n_\gamma}\left(\hat{\theta}_\gamma-\theta_0\right)$ is asymptotically equivalent to $\left(\sqrt{n_\gamma}\bar{\mathscr{L}}\left(\left(Y_{\gamma\varsigma_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right)-\sqrt{n_\gamma}\left(\hat{\xi}_\gamma-\xi_0\right)\mathscr{I}_{12}\right)/\mathscr{I}_{11}$. This variable converges in distribution to $\mathscr{N}\left(0,\sigma^2\right)$, establishing then Theorem 5.2.

D.4 Proofs for stratified sampling

D.4.1 Proof of Result 5.1

Proof. We first show that A5.0 is satisfied.

- A5.0a is satisfied: $M_{\theta,\xi}:y\mapsto=\sup_\gamma\{n_\gamma^{-1}N_\gamma\}$, is convenient: $M_{\theta,\xi}<+\infty$ because $\lim_{\gamma\to\infty}\{n_\gamma^{-1}N_\gamma\}<+\infty$.
- To show that A5.0b is satisfied, we calculate:

$$\begin{aligned}\lim_\gamma P_{\theta,\xi}(I_{\gamma k}=1|Y_k=y,Z_k=z)&=\lim_\gamma P_{\theta,\xi}(I_{\gamma k}=1|Z_k)\\ &=\sum_{h=1}^{H}\tau_{\infty h}\mathbb{1}_{\left[\Phi^{-1}\left(t_{\infty,h-1}\right),\Phi^{-1}\left(t_{\infty,h}\right)\right]}\left(\frac{Z_k-\theta\xi}{\sqrt{\xi^2+\sigma^2}}\right).\end{aligned}$$

We deduce from the preceding that $m_{\infty,\theta,\xi}(y)$ is defined:

$$\begin{aligned}&m_{\infty,\theta,\xi}(y)\\ &=\sum_{h=1}^{H}\tau_{\infty h}\,\mathrm{P}_{\theta,\xi}\left(\xi\theta+\sqrt{\xi^2+\sigma^2}\Phi^{-1}\left(t_{\infty,h}\right)<\xi y+\varepsilon<\xi\theta+\sqrt{\xi^2+\sigma^2}\Phi^{-1}\left(t_{\infty,h-1}\right)\right)\\ &=\sum_{h=1}^{H}\tau_{\infty h}\left(\Phi\left(\sqrt{\xi^2+\sigma^2}\left(\Phi^{-1}\left(t_{\infty,h}\right)\right)+\xi(\theta-y)\right)-\Phi\left(\sqrt{\xi^2+\sigma^2}\Phi^{-1}\left(t_{\infty,h-1}\right)+\xi(\theta-y)\right)\right).\end{aligned}$$

By assumption, $\int m_{\infty,\theta,\xi}f_\theta d\lambda=\lim_{\gamma\to\infty}{N_\gamma}^{-1}n_\gamma>0$, so $\rho_{\infty,\theta,\xi}$ is defined:

$$\begin{aligned}&\rho_{\infty,\theta,\xi}(y)\\ &=\frac{\sum_{h=1}^{H}\tau_{\infty h}\left(\Phi\left(\sqrt{\xi^2+\sigma^2}\left(\Phi^{-1}(t_{\infty,h})\right)+\xi(\theta-y)\right)-\Phi\left(\sqrt{\xi^2+\sigma^2}\Phi^{-1}(t_{\infty,h-1})+\xi(\theta-y)\right)\right)}{\left(\sum_{h=1}^{H}\tau_{\infty h}\left(t_{\infty,h}-t_{\infty,h-1}\right)\right)}.\end{aligned}$$

□

D.4.2 Proof of Result 5.3

Proof.

- We show that A5.2 is satisfied: A5.2a is satisfied because
  - A5.1a is satisfied: $\int M_{\theta,\xi} f_\theta d\lambda_{\mathbb{R}^p} < +\infty$.
  - A5.1b is satisfied:

$$
\begin{aligned}
& c_{\gamma,\theta,\xi}(y_1, y_2) \\
&= \sum_{h=1}^{H}\sum_{h'=1}^{H} \tau_{\gamma h}\tau_{\gamma h'} \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \cap \left\{Z_{(T_{\gamma h'-1})} < Z_2 \leqslant Z_{(T_{\gamma h'})}\right\} \middle| Y_1, Y_2\right) \\
&\quad - \left(\sum_{h=1}^{H} \tau_{\gamma h}\, \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right)\right) \\
&\qquad \left(\sum_{h=1}^{H} \tau_{\gamma h}\, \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_2 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right)\right) \\
&\quad - \sum_{h=1}^{H} \frac{n_{\gamma h}\,(N_{\gamma h} - n_{\gamma h})}{N_{\gamma h}^2\,(N_{\gamma h} - 1)} \\
&\qquad \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \cap \left\{Z_{(T_{\gamma h-1})} < Z_2 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right).
\end{aligned}
$$

    Besides, $\forall h \in \{1, \ldots, H\}$:

$$
\begin{aligned}
& \left|\sum_{h=1}^{H} \frac{n_{\gamma h}(N_{\gamma h} - n_{\gamma h})}{N_{\gamma h}^2\,(N_{\gamma h} - 1)}\, \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \cap \left\{Z_{(T_{\gamma h-1})} < Z_2 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right)\right| \\
&\quad \leqslant \sum_{h=1}^{H} \frac{n_{\gamma h}\,(N_{\gamma h} - n_{\gamma h})}{N_{\gamma h}^2\,(N_{\gamma h} - 1)} = O_\gamma(N_\gamma^{-1}),
\end{aligned}
$$

    and it is possible to show that $\forall h, h'$

$$
\begin{aligned}
& \Big|\mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \cap \left\{Z_{(T_{\gamma h'-1})} < Z_2 \leqslant Z_{(T_{\gamma h'})}\right\} \middle| Y_1, Y_2\right) - \\
&\qquad \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right) \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h'-1})} < Z_2 \leqslant Z_{(T_{\gamma h'})}\right\} \middle| Y_1, Y_2\right)\Big| \\
&= \Big|\mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\} \middle| Y_1, Y_2\right)\left(\mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h'-1})} < Z_2 \leqslant Z_{(T_{\gamma h'})}\right\} \middle| Y_1, Y_2\right)\right. \\
&\qquad \left. - \mathrm{P}_{\theta_0,\xi_0}\left(\left\{Z_{(T_{\gamma h'-1})} < Z_2 \leqslant Z_{(T_{\gamma h'})}\right\} \middle| \left\{Z_{(T_{\gamma h-1})} < Z_1 \leqslant Z_{(T_{\gamma h})}\right\}, Y_1, Y_2\right)\right)\Big| \\
&= o_\gamma(N_\gamma^{-1}).;
\end{aligned}
$$

  - A5.1c is satisfied: the proof is quite similar to the proof of A5.1b.
  - A5.1d is satisfied: we are in the case of sampling without replacement, and we can show that $v_{\gamma,\theta,\xi} + m^2_{\gamma,\theta,\xi} < 1 = o_\gamma(N_\gamma)$
  - A5.1e is satisfied: the probability to have an empty sample is 0.
- A5.2b is satisfied because $F \in \mathscr{C}^\infty(\mathbb{R}, [0,1])$, so Δ is a composition of infinitely differentiable functions.

- A5.2c is satisfied. Calculations are omitted.

- In order to verify A5.2d, we just have to observe that:

$$\begin{aligned}\mathscr{I}_{11} &< \tau^{-1}\max_{h=1}^{H}\{\tau_{\infty h}\}\int_{\mathbb{R}^p}\left(\frac{\partial\Delta}{\partial\theta}(y,\theta_0,\xi_0)\right)^2 d\lambda_{\mathbb{R}^p}(y)\\ &< \tau^{-1}\max_{h=1}^{H}\{\tau_{\infty h}\}\int_{\mathbb{R}^p}\left(\frac{\partial\Delta}{\partial\theta}(y,\theta_0,\xi_0)\right)^2 d\lambda_{\mathbb{R}^p}(y)\\ &< +\infty,\end{aligned}$$

because

$$\begin{aligned}&\left(\frac{\partial}{\partial\theta}\Delta\right)(Y_1,\theta_0,\xi_0)\\ &= (y-\theta)+\frac{\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\frac{\xi}{\sigma}f_0\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2+1}\,\Phi^{-1}(t_{\infty,h})+\frac{\xi}{\sigma}(\theta-y)\right)}{\tau_H+\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\,\Phi\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2+1}\,\Phi^{-1}(t_{\infty,h})+\frac{\xi}{\sigma}(\theta-y)\right)}\\ &\leqslant (y-\theta)+\frac{\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\frac{\xi}{\sigma}\frac{1}{\sqrt{2\pi}}}{\min\{\tau_{\infty h}\}}.\end{aligned}$$

We have

$$\begin{aligned}&\left(\frac{\partial}{\partial\xi}\Delta\right)(Y_1,\theta_0,\xi_0)\\ &= \frac{\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\left(\frac{\xi/\sigma^2}{\sqrt{\frac{\xi^2}{\sigma^2}+1}}\Phi^{-1}(t_{\infty,h})+\frac{(\theta-y)}{\sigma}\right)f_0\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2+1}\,\Phi^{-1}(t_{\infty,h})+\frac{\xi}{\sigma}(\theta-y)\right)}{\tau_H+\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\,\Phi\left(\sqrt{\left(\frac{\xi}{\sigma}\right)^2+1}\,\Phi^{-1}(t_{\infty,h})+\frac{\xi}{\sigma}(\theta-y)\right)}\\ &\leqslant \frac{\sum_{h=1}^{H-1}(\tau_{\infty h}-\tau_{\infty,h+1})\left(\frac{\xi/\sigma^2}{\sqrt{\frac{\xi^2}{\sigma^2}+1}}+\frac{(\theta-y)}{\sigma}\right)\frac{1}{\sqrt{2\pi}}}{\min\{\tau_{\infty h}\}}.\end{aligned}$$

$\left|\frac{\partial\Delta}{\partial\theta}\frac{\partial\Delta}{\partial\xi}(y,\theta_0,\xi_0)\right|\rho_{\infty,\theta_0,\xi_0}$ can be upper bounded by a function of the form (ay^2+by+c) so $\mathrm{E}_{\theta_0,\xi_0}\left[\left|\left(\frac{\partial}{\partial\theta}\Delta\right)\left(\frac{\partial}{\partial\psi}\Delta\right)\right|\right] < +\infty$ and $\mathscr{I}_{12}$ is defined.

- A5.2e is satisfied:

$$\begin{aligned}
&\left|\frac{\partial}{\partial\xi}\ln\left(\frac{\rho_{\infty,\theta,\xi}(y)f_\theta(y)}{\rho_{\infty\theta_0\xi}(y)f_{\theta_0}(y)}\right)\right| \\
&= \left|\ln\left(\frac{\tau_H+\sum_{h=1}^H(\tau_{\infty h}-\tau_{\infty h+1})\,\Phi\left(\sqrt{\xi_0^2+\sigma^2}\left(\Phi^{-1}(t_{\infty,h})\right)+\xi_0(\theta_0-y)\right)}{\tau_H+\sum_{h=1}^H(\tau_{\infty h}-\tau_{\infty h+1})\,\Phi\left(\sqrt{\xi_0^2+\sigma^2}\left(\Phi^{-1}(t_{\infty,h})\right)+\xi_0(\theta_0-y)\right)}\right)\right. \\
&\qquad \left.+(\theta-\theta_0)\right| \\
&< K(y,\theta) = \left|\ln\left(\frac{\max_h\{\tau_{\infty h}\}}{\min_{h=1}^H\{\tau_{\infty h}\}}\right)+(\theta-\theta_0)\right|
\end{aligned}$$

The function $K(.,\theta)$ is constant, so, as $\mathbb{1}$ satisfies A5.1, $K(.,\theta)$ satisfies A5.1.

- A5.2f is satisfied: $\exists L:\mathbb{R}^p\times A\to\mathbb{R}^+$ a measurable function such that $\forall\theta\in A,\xi\in B,\lambda_{\mathbb{R}^p}-a.s(y)$,

$$\left|\frac{\partial^3\Delta}{\partial\theta^3}(y,\theta,\xi)\right|\leqslant L(y),\qquad \left|\frac{\partial^3\Delta}{\partial\theta^2\partial\xi}(y,\theta_0,\xi)\right|\leqslant L(y),\qquad \left|\frac{\partial^3\Delta}{\partial\theta\partial\xi^2}(y,\theta_0,\xi)\right|\leqslant L(y)$$

$$\text{and}\quad\int L(y)\rho_{\infty,\theta_0,\xi_0}(y)f_{\theta_0}(y)\,\mathrm{d}\lambda_{\mathbb{R}^p}(y)<\infty. \tag{D.10}$$

- A5.2g is satisfied, proof is omitted.
- A5.2h is satisfied, proof is omitted.

□

D.4.3 Asymptotic normality

Lemma D.2. *Consider the asymptotic framework described in section 5.1.1. Consider stratified sampling: Let $H\in\mathbb{N}\backslash\{0\}$ be a fixed and non-random number of strata, and let $(N_{\gamma h})_{\gamma\in\mathbb{N},h\in\{1,\dots,H\}}$ be an array of strictly positive integers such that $\forall\gamma\in\mathbb{N}$, $N_\gamma=\sum_{h=1}^H N_{\gamma h}$. For $\gamma\in\mathbb{N}$, $h\in\{1,\dots,H\}$, let $n_{\gamma h}\in\{1,\dots,N_{\gamma h}\}$. We define ν_γ the permutation such that $Z_{\nu_\gamma(1)}<\dots<Z_{\nu_\gamma(N_\gamma)}$. The permutation ν_γ is a random variable which is a function of $\mathcal{Z}_\gamma$. The hth stratum of the γth population is the set: $U_{\gamma h}=\nu_\gamma\left(\{T_{\gamma h-1}+1,\dots,T_{\gamma h}\}\right)$, with $T_{\gamma 0}=0$, $T_{\gamma h}=\sum_{1\leqslant h'\leqslant,h}N_{\gamma h'}$. For $h\in\{0,\dots,H\}$, define $t_{\gamma h}=\frac{T_{\gamma h}}{N_\gamma}$. The random design measure is stratified simple random sampling without replacement:*

$$\Pi_\gamma(\{i\})=\begin{vmatrix}\prod_{h=1}^H\binom{N_{\gamma h}}{n_{\gamma h}}^{-1} & \text{if } \forall h\in\{1,\dots,H\},\sum_{k\in U_{\gamma h}}i_k=n_{\gamma h},\text{ and } i\in\{0,1\}^{N_\gamma}\\ 0 & \text{otherwise.}\end{vmatrix}$$

For $h\in\{0,\dots,H\}$, we define $t_{\gamma h}=\frac{T_{\gamma h}}{N_\gamma}$. We assume that $\forall h\in\{0,\dots,H\}$, $t_{\infty,h}=\lim_{\gamma\to\infty}t_{\gamma h}$ is defined, and that $\forall h\in\{0,\dots,H\}$, $t_{\infty,h-1}<t_{\infty h}$. We also define $\tau_{\gamma h}=N_{\gamma h}^{-1}n_{\gamma h}$ and we assume that $\forall h\in\{1,\dots,H\}$, $\tau_{\infty h}=\lim_{\gamma\to\infty}\tau_{\gamma h}$ is defined and strictly positive. Together with the fact that $\forall\gamma,h$, $n_{\gamma h}>0$, this implies that $m=\min\{\tau_{\gamma h}|\gamma\in\mathbb{N},h\in\{1,\dots,H\}\}>0$. We assume that $\tau=\sum_{h=1}^H\tau_{\infty h}(t_{\infty,h}-t_{\infty,h-1})=\lim_{\gamma\to\infty}N_\gamma^{-1}n_\gamma^*>0$, with $n_\gamma^*=\sum_{k=1}^H n_{\gamma h}$.

Let $p\in\mathbb{N}\backslash\{0\}$. We define a sequence of random vectors $(\mathcal{X}_\gamma)_{\gamma\in\mathbb{N}}$. For $\gamma\in\mathbb{N}$, $\mathcal{X}_\gamma=\left(X_{\gamma1}\dots X_{\gamma N_\gamma}\right)$, where for $k\in\{1,\dots,N_\gamma\}$, $X_{\gamma k}$ is with value in $\mathbb{R}^p$. We assume that there exists $g:\left(\mathscr{Y}\times\mathscr{Z}\times[0,1],\mathscr{T}_Y\otimes\mathscr{T}_Z\otimes\mathscr{B}_{[0,1]}\right)\to(\mathbb{R}^p,\mathscr{B}_{\mathbb{R}^p})$ a measurable function such that $\forall\gamma\in\mathbb{N}$, $\forall k\in\{1,\dots,N_\gamma\}$, $X_{\gamma k}=g(Y_k,Z_k,\pi_{\gamma k})$. We assume that there exists $G:(\mathscr{Y}\times\mathscr{Z},\mathscr{T}_Y\otimes\mathscr{T}_Z)\to(\mathbb{R},\mathscr{B}_{\mathbb{R}})$

such that $\mathrm{E}\left[G(Y_1,Z_1)^2\right] < +\infty$, *and* $\forall y,z,\pi \in \mathscr{Y}\times\mathscr{Z}\times[m,1]$, $\|g(y,z,\pi)\| \leqslant G(y,z)$. *We assume that* $\forall y \in \mathscr{Y}$, $g(y,.,.) : \mathscr{Z}\times]0,1] \to \mathbb{R}, (z,\pi) \mapsto g(y,z,\pi)$ *is continuous. Then, we define* $S_{\gamma h} = \sum_{k\in U_{\gamma h}} X_{\gamma k} I_{\gamma k}$ *and* $S_\gamma = \sum_{h=1}^H S_{\gamma h}$. *We also define*

$$\begin{aligned}
V_{\infty,h} &= \mathrm{Var}\left[g\left(Y_k,Z_k,\tau_{\infty h}\right)\middle|\, Z_k \in \left]\zeta(t_{\infty h-1}),\zeta(t_{\infty h})\right]\right] \\
E_{\infty,h} &= \mathrm{E}\left[g\left(Y_k,Z_k,\tau_{\infty h}\right)\middle|\, Z_k \in \left]\zeta(t_{\infty h-1}),\zeta(t_{\infty h})\right]\right] \\
V_\infty &= \tau^{-1}\sum_{h=1}^H \left(t_{\infty h}-t_{\infty h-1}\right)\tau_{\infty h}V_{\infty h} \\
E_\infty &= \sum_{h=1}^H \tau_{\infty h}E_{\infty h}
\end{aligned}$$

We assume that the quantile function of Z_1 *is continuous on* $]0,1[$, *and that* $\forall z_0 \in \mathbb{R}$, $\exists O$ *an open subset of* $\mathbb{R}$, $\exists M : (\mathscr{Y},\mathscr{T}_Y) \to (\mathbb{R},\mathscr{B}_\mathbb{R})$ *a positive and measurable function such that* $\int M(y)\,\mathrm{d}\lambda < +\infty$, $\int \sup_{z\in O}\{G(y,z)\}M(y)\,\mathrm{d}\lambda(y) < +\infty$, $\int \sup_{z\in O}\{G^2(y,z)\}M(y)\,\mathrm{d}\lambda(y) < +\infty$ *and* $\forall z \in K, y \in \mathscr{Y}$, $\frac{\mathrm{dP}^{Y_1|Z_1=z}}{\mathrm{d}\lambda}(y) < M(y)$. *We assume that* $\forall y \in \mathscr{Y}, z \mapsto \left(\mathrm{dP}^{Y_1|Z_1=z}/\,\mathrm{d}\lambda\right)(y)$ *is continuous. Then*

$$\sqrt{n_\gamma}\left(n_\gamma^{-1}S_\gamma - E_\infty\right) \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}(0,V_\infty)$$

Proof. Let $\zeta : y \mapsto \inf\{x|P(Z_1 \leqslant x) \geqslant y\}$ be the quantile function of Z_1 (see Serfling (1980, p. 74)). The continuity of ζ ensures the strong consistency of the sample quantiles (see Serfling (1980, p. 75)), and further: $\mathrm{P}\left(\bigcap_{h=1}^{H-1}\left\{\omega\in\Omega\middle|\lim_{\gamma\to\infty} Z_{\nu_\gamma(T_{\gamma,h})}(\omega) = \zeta(t_{\infty,h})\right\}\right) = 1$. Let $t^*_{\gamma,h}$ be a sequence defined for all $\gamma \in \mathbb{N}$, $h \in \{1,\ldots,H\}$, such that $h < h' \Rightarrow 0 < t^*_{\gamma,h} \leqslant t^*_{\gamma,h'} \leqslant 1$, and such that $\forall h \in \{1,\ldots,H\}$, $\lim_{\gamma\to\infty} t^*_{\gamma,h} = t_{\infty,h}$. Then conditionally on $Z_{\nu_\gamma(T_{\gamma,h-1})} = \zeta\left(t^*_{\gamma,h-1}\right), Z_{\nu_\gamma(T_{\gamma,h})} = \zeta\left(t^*_{\gamma,h}\right)$, we have independence within the same stratum: $\forall h \in \{1,\ldots,H-1\}$,

$$\begin{aligned}
&P^{(Z_k,Y_k)_{k\in r_{\gamma h}\circ\nu_\gamma\left(U_{\gamma,h}\right)}|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \\
&= \left(P^{(Z_1,Y_1)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)^{\otimes N_{\gamma,h}-1} \otimes P^{(Z_1,Y_1)|Z_1=\zeta\left(t^*_{\gamma,h}\right)},
\end{aligned} \tag{D.11}$$

and

$$P^{(Z_k,Y_{\gamma k})_{k\in r_{\gamma H}\circ\nu_\gamma\left(U_{\gamma,H}\right)}|Z_{\nu_\gamma(T_{\gamma,H-1})}=\zeta\left(t^*_{\gamma,H-1}\right)} = \left(P^{(Z_1,Y_1)|Z_1\in\left]\zeta(t^*_{\gamma,H-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)^{\otimes N_{\gamma,H}}, \tag{D.12}$$

where for $h \in \{1,\ldots,H-1\}$, $r_{\gamma h}$ is a random permutation of the ordered set $\nu_\gamma\left(U_{\gamma,h}\right)$ such that $r_{\gamma h}\circ\nu_\gamma\left(T_{\gamma,h}\right) = T_{\gamma,h}$, and $r_{\gamma H}$ is a random permutation of the ordered set $\nu_\gamma\left(U_{\gamma,H}\right)$. If we consider the distribution of the sample responses on hth stratum, we have: $\forall h \in \{1,\ldots,H-1\}$,

$$\begin{aligned}
&P^{(Z_k,Y_{\gamma k})_{k\in r_{\gamma h}\circ\nu_\gamma\left(U_{\gamma,h}\right),I_{\gamma k}=1}|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \\
&= \left(P^{(Z_1,Y_1)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)^{\otimes n_{\gamma,h}-1} \\
&\quad\otimes\left(\tau_{\gamma h}P^{(Z_1,Y_1)|Z_1=\zeta\left(t^*_{\gamma,h}\right)} + (1-\tau_{\gamma h})\left(P^{(Z_1,Y_1)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)\right),
\end{aligned} \tag{D.13}$$

and

$$P^{(Z_k,Y_{\gamma k})_{k\in r_{\gamma H}\circ\nu_\gamma\left(U_{\gamma,H}\right),I_{\gamma k}=1}|Z_{\nu_\gamma(T_{\gamma,H-1})}=\zeta\left(t^*_{\gamma,H-1}\right)} = \left(P^{(Z_1,Y_1)|Z_1\in\left]\zeta(t^*_{\gamma,H-1}),+\infty\right]}\right)^{\otimes n_{\gamma,H}}. \tag{D.14}$$

Equation (D.13) implies that $\forall h \in \{1, \ldots, H-1\}$,

$$\begin{aligned} & P^{(X_{\gamma k})_{k\in\nu_\gamma(U_{\gamma,h}),I_{\gamma k}=1}|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \\ = \;& \tau_{\gamma h} P^{g\left(Y_1,Z_1,\tau_{\gamma h}\right)|Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \otimes \left(P^{g\left(Y_1,Z_1,\tau_{\gamma h}\right)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)^{\otimes n_{\gamma,h}-1} \\ & + (1-\tau_{\gamma h}) \left(P^{g\left(Y_1,Z_1,\tau_{\gamma h}\right)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}\right)^{\otimes n_{\gamma,h}}, \end{aligned} \tag{D.15}$$

and equation (D.14) implies that

$$P^{(X_{\gamma k})_{k\in\nu_\gamma(U_{\gamma,H}),I_{\gamma k}=1}|Z_{\nu_\gamma(T_{\gamma,H-1})}=\zeta\left(t^*_{\gamma,H-1}\right)} = \left(P^{g\left(Y_1,Z_1,\tau_{\gamma H}\right)|Z_1\in\left]\zeta(t^*_{\gamma,h-1}),+\infty\right]}\right)^{\otimes n_{\gamma,H}}. \tag{D.16}$$

We will show that

$$\forall h \in \{0, \ldots, H\}\,,\; \mathrm{P}^{\sqrt{n_{\gamma h}}\left(n_{\gamma h}^{-1}S_{\gamma,h}-E_{\gamma h}\right)|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),\, Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}(0, V_{\infty h}). \tag{D.17}$$

Using (D.15) we calculate, for $h \in \{0, \ldots, H-1\}$:

$$\begin{aligned} & \mathrm{Var}\left[S_{\gamma h}|Z_{\nu_\gamma(T_{\gamma,h-1})} = \zeta\left(t^*_{\gamma,h-1}\right),\, Z_{\nu_\gamma(T_{\gamma,h})} = \zeta\left(t^*_{\gamma,h}\right]\right) \\ = \;& \left(n^*_\gamma - 1\right) \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ & + \tau_{\gamma h}\, \mathrm{E}\left[\left(g\, g^{\mathrm{T}}\right)\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 = \zeta\left(t^*_{\gamma,h}\right)\right] \\ & \quad + (1-\tau_{\gamma h})\, \mathrm{E}\left[\left(g\, g^{\mathrm{T}}\right)\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ & - \left(\tau_{\gamma h}\, \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 = \zeta\left(t^*_{\gamma,h}\right)\right]\right. \\ & \quad \left. + (1-\tau_{\gamma h})\, \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right]\right) \\ & \left(\tau_{\gamma h}\, \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 = \zeta\left(t^*_{\gamma,h}\right)\right]\right. \\ & \quad \left. + (1-\tau_{\gamma h})\, \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right]\right)^{\mathrm{T}}, \end{aligned} \tag{D.18}$$

and

$$\mathrm{Var}\left[S_{\gamma H}|Z_{\nu_\gamma(T_{\gamma,H-1})} = \zeta\left(t^*_{\gamma,H-1}\right)\right] = n^*_\gamma\, \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), +\infty\right]\right]. \tag{D.19}$$

In addition,

$$\begin{aligned} & \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ = \;& \left(\mathrm{P}\left(Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right)\right)^{-1} \int g\left(y, z, \tau_{\gamma h}\right) \mathbb{1}_{\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}(z)\, \mathrm{dP}^{Y_1,Z_1}(y, z), \end{aligned}$$

and

$$\begin{aligned} & \mathrm{E}\left[\left(g\, g^{\mathrm{T}}\right)\left(Y_1, Z_1, \tau_{\gamma h}\right)|\, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ = \;& \left(\mathrm{P}\left(Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right)\right)^{-1} \int \left(g\, g^{\mathrm{T}}\right)\left(y, z, \tau_{\gamma h}\right) \mathbb{1}_{\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}(z)\, \mathrm{dP}^{Y_1,Z_1}(y, z). \end{aligned}$$

Because $\forall (y, z) \in \mathscr{Y} \times \mathscr{Z}$,

$$\lim_{\gamma\to\infty} g(y, z, \tau_{\gamma h}) \mathbb{1}_{\left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]}(z) = g(y, z, \tau_{\infty h}) \mathbb{1}_{\left]\zeta(t_{\infty,h-1}),\zeta\left(t_{\infty,h}\right)\right]}(z),$$

and $\left\| g(y,z,\tau_{\gamma h}) \mathbb{1}_{]\zeta(t^*_{\gamma,h-1}),\zeta(t^*_{\gamma,h})]}(z) \right\| \leqslant G(y,z)$, we conclude by the Lebesgue dominated convergence theorem that

$$\begin{aligned} &\lim_{\gamma\to\infty} \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right) | \, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ &\quad = \quad \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 \in \left]\zeta(t_{\infty,h-1}), \zeta\left(t_{\infty,h}\right)\right]\right], \qquad \text{(D.20)} \\ &\lim_{\gamma\to\infty} \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right) | \, Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right] \\ &\quad = \quad \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 \in \left]\zeta(t_{\infty,h-1}), \zeta\left(t_{\infty,h}\right)\right]\right]. \qquad \text{(D.21)} \end{aligned}$$

Also, as $\forall y \in \mathscr{Y} \times \mathscr{Z}$,

$$\lim_{\gamma\to\infty} g\left(y, \zeta\left(t^*_{\gamma,h}\right), \tau_{\gamma h}\right) \frac{\mathrm{dP}^{Y_1|Z_1=\zeta\left(t^*_{\gamma,h}\right)}}{\mathrm{d}\lambda}(y) = g\left(y, \zeta\left(t^*_{\infty,h}\right), \tau_{\gamma h}\right) \frac{\mathrm{dP}^{Y_1|Z_1=\zeta\left(t^*_{\infty,h}\right)}}{\mathrm{d}\lambda}(y)$$

and for γ large enough, $\left\| g\left(y, \zeta\left(t^*_{\gamma,h}\right), \tau_{\gamma h}\right) \frac{\mathrm{dP}^{Y_1|Z_1=\zeta\left(t^*_{\gamma,h}\right)}}{\mathrm{d}\lambda}(y) \right\| \leqslant G(y,z)M(y)$, we conclude by the Lebesgue dominated convergence theorem that:

$$\begin{aligned} \lim_{\gamma\to\infty} \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right) | \, Z_1 = \zeta(t^*_{\gamma,h})\right] &= \mathrm{E}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 = \zeta\left(t_{\infty,h}\right)\right] \quad < +\infty, \qquad \text{(D.22)} \\ \lim_{\gamma\to\infty} \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\gamma h}\right) | \, Z_1 = \zeta(t^*_{\gamma,h})\right] &= \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 = \zeta\left(t_{\infty,h}\right)\right] \quad < +\infty. \qquad \text{(D.23)} \end{aligned}$$

Equations (D.18), (D.19) and the preceding imply that

$$\begin{aligned} &\lim_{\gamma\to\infty} \left(n_{\gamma h}\right)^{-1} \mathrm{Var}\left[S_{\gamma h} | Z_{\nu_\gamma(T_{\gamma,h-1})} = \zeta\left(t^*_{\gamma,h-1}\right), \, Z_{\nu_\gamma(T_{\gamma,h})} = \zeta\left(t^*_{\gamma,h}\right)\right] \\ &\quad = \quad \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 \in \left]\zeta(t_{\infty,h-1}), \zeta\left(t_{\infty,h}\right)\right]\right] \end{aligned}$$

and

$$\begin{aligned} &\lim_{\gamma\to\infty} \left(n_{\gamma H}\right)^{-1} \mathrm{Var}\left[S_{\gamma H} | Z_{\nu_\gamma(T_{\gamma,H-1})} = \zeta\left(t^*_{\gamma,h-1}\right)\right] \\ &\quad = \quad \mathrm{Var}\left[g\left(Y_1, Z_1, \tau_{\infty h}\right) | \, Z_1 \in \left]\zeta(t_{\infty,H-1}), +\infty\right]\right]. \end{aligned}$$

For $\gamma \in \mathbb{N}, h \in \{1, \ldots, H-1\}$ introduce the random variables $X^*_{\gamma,h,1} \ldots X^*_{\gamma,h,n_{\gamma h}}$ that satisfy

$$\begin{aligned} \mathrm{P}^{X^*_{\gamma,h,1} \cdots X^*_{\gamma,h,n_{\gamma h}}} \quad = \quad & \left(P^{g(Z_1,Y_1,\tau_{\gamma h})|Z_1 \in \left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]} \right)^{\otimes n_{\gamma,h}-1} \\ & \otimes \left(\tau_{\gamma h} P^{g(Z_1,Y_1,\tau_{\gamma h})|Z_1=\zeta\left(t^*_{\gamma,h}\right)} + (1-\tau_{\gamma h}) \left(P^{g(Z_1,Y_1,\tau_{\gamma h})|Z_1 \in \left]\zeta(t^*_{\gamma,h-1}),\zeta\left(t^*_{\gamma,h}\right)\right]} \right) \right), \end{aligned}$$

and the random variables $X^*_{\gamma,H,1} \ldots X^*_{\gamma,H,n_{\gamma H}}$ that satisfy

$$\mathrm{P}^{X^*_{\gamma,H,1} \cdots X^*_{\gamma,H,n_{\gamma H}}} \quad = \quad \left(P^{g(Z_1,Y_1,\tau_{\gamma H})|Z_1 > \zeta(t^*_{\gamma,H-1})} \right)^{\otimes n_{\gamma H}}.$$

For $\alpha \in \mathbb{R}^p$, $h \in \{1, \ldots, H\}$,

$$\mathrm{P}^{\alpha^{\mathrm{T}} S_{\gamma h}} = \mathrm{P}^{\sum_{k=1}^{n_{\gamma h}} \alpha^{\mathrm{T}} S_{\gamma h}}$$

Where $^{\mathrm{T}}$ is the transpose operator. For $\alpha \in \mathbb{R}^p \backslash \{0\}$, $\gamma\varepsilon \in]0, +\infty]$, we define

$$A_{\gamma,h,\varepsilon,\alpha} = \frac{\sum_{k=1}^{n_{\gamma h}} \mathrm{E}\left[\left| \alpha^{\mathrm{T}} \left(X^*_{\gamma,h,k} - \mathrm{E}\left[X^*_{\gamma,h,k} \right] \right) \right|^2 \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \mathrm{Var}[S_\gamma]\alpha}, +\infty\right]} \left(\left| \alpha^{\mathrm{T}} \left(X^*_{\gamma,h,k} - \mathrm{E}\left[X^*_{\gamma,h,k} \right] \right) \right| \right) \right]}{\alpha^{\mathrm{T}} \mathrm{Var}\left[S_{\gamma h} \right] \alpha}.$$

Let $\alpha \in \mathbb{R}^p \backslash \{0\}$. To prove the asymptotic normality of $\alpha^{\mathrm{T}} S_{\gamma h}$, we will show that the Lindeberg condition

$$\forall \varepsilon \in]0, +\infty[\,, \lim_{\gamma \to \infty} A_{\gamma,h,\varepsilon,\alpha} = 0 \tag{D.24}$$

is satisfied. Let $\varepsilon \in]0, +\infty[$, $h \in \{1, \ldots, H\}$. Then

$$\begin{aligned} & A_{\gamma,h,\varepsilon,\alpha} \\ & \sim_\gamma \frac{(n_{\gamma h} - 1)}{n_{\gamma h}} \frac{\mathrm{E}\left[\left|\, \alpha^{\mathrm{T}}\Big(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma hk}\right]\Big)\right|^2 \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}[S_\gamma]\alpha}, +\infty\right]} \left(\alpha^{\mathrm{T}} \left(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma hk}\right]\right)\right)\right]}{\alpha^{\mathrm{T}} V_{\infty h} \alpha}. \end{aligned} \tag{D.25}$$

In addition,

$$\begin{aligned} & \int \left|\, \alpha^{\mathrm{T}} \left(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma h1}\right]\right)\right|^2 \\ & \qquad \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}[S_\gamma]\alpha}, +\infty\right]} \left(\alpha^{\mathrm{T}} \left(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma h1}\right]\right)\right) \mathrm{dP}^{Y_1, Z_1} \\ \leqslant & \int |\alpha\|^2 \ \left(\|G(Y_1, Z_1)\| + \left\|\mathrm{E}\left[X^*_{\gamma h1}\right]\right\|\right)^2 \\ & \qquad \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}[S_\gamma]\alpha}, +\infty\right]} \left(|\alpha\|^2 \ \left(\|G(Y_1, Z_1)\| + \left\|\mathrm{E}\left[X^*_{\gamma h1}\right]\right\|\right)\right) \mathrm{dP}^{Y_1, Z_1} \\ \leqslant & \int |\alpha\|^2 \ \left(\|G(Y_1, Z_1)\| + \frac{\mathrm{E}\left[G(Y_1, Z_1)\right]}{\min\{t^*_{\gamma h} - t^*_{\gamma,h-1} | \gamma \in \mathbb{N}\}}\right)^2 \\ & \qquad \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}[S_\gamma]\alpha}, +\infty\right]} \left(\|\alpha\| \ \left(\|G(Y_1, Z_1)\| + \frac{\mathrm{E}\left[G(Y_1, Z_1)\right]}{\min\{t^*_{\gamma h} - t^*_{\gamma,h-1} | \gamma \in \mathbb{N}\}}\right)\right) \mathrm{dP}^{Y_1, Z_1} \end{aligned}$$

because, for $h \in \{1, \ldots, H\}$, with the convention $t^*_{\gamma H} = 1$,

$$\begin{aligned} \left\|\mathrm{E}\left[X^*_{\gamma h1}\right]\right\| & \leqslant \frac{\int G(Y_1, Z_1) \, \mathrm{dP}^{Y_1, Z_1}}{\mathrm{P}\left(Z_1 \in \left]\zeta(t^*_{\gamma,h-1}), \zeta\left(t^*_{\gamma,h}\right)\right]\right)} \\ & \leqslant \frac{\int G(Y_1, Z_1) \, \mathrm{dP}^{Y_1, Z_1}}{t^*_{\gamma h} - t^*_{\gamma,h-1}} \\ & \leqslant \frac{\int G(Y_1, Z_1) \, \mathrm{dP}^{Y_1, Z_1}}{\min\{t^*_{\gamma h} - t^*_{\gamma,h-1} | \gamma \in \mathbb{N}\}}, \end{aligned}$$

because $\mathrm{P}^{(T^*_{\gamma h}, T^*_{\gamma,h-1})_{\gamma \in \mathbb{N}}} - a.s. \left((t^*_{\gamma h} - t^*_{\gamma,h-1})_{\gamma \in \mathbb{N}}\right)$, $\lim_\gamma t^*_{\gamma h} - t^*_{\gamma,h-1} = t^*_{\infty h} - t^*_{\infty,h-1}$ and because $\min\{t^*_{\gamma h} - t^*_{\gamma,h-1} | \gamma \in \mathbb{N}\} > 0$. As $\lim_{\gamma \to \infty} \varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}\left[S_\gamma\right] \alpha} = +\infty$, and $\mathrm{E}\left[G(Y_1, Z_1)^2\right] < +\infty$, we conclude that

$$\lim_{\gamma \to \infty} \int \left|\, \alpha^{\mathrm{T}} \left(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma h1}\right]\right)\right|^2 \mathbb{1}_{\left]\varepsilon\sqrt{\alpha^{\mathrm{T}} \operatorname{Var}[S_\gamma]\alpha}, +\infty\right]} \left(\alpha^{\mathrm{T}} \left(g(Y_1, Z_1, \tau_{\gamma h}) - \mathrm{E}\left[X^*_{\gamma h1}\right]\right)\right) = 0, \tag{D.26}$$

which implies via (D.25) that the Lindeberg condition (D.24) is satisfied. We apply the Lindeberg-Feller theorem (see (Serfling, 1980, Theorem p. 31)), and conclude by the asymptotic normality of $\alpha^{\mathrm{T}} S_{\gamma h}$ conditionally on $(T_{\gamma h})$ $\forall \alpha \in \mathbb{R}^p$ (which terminates the proof of (D.17)).

Then, we remark that conditionally on $Z_{\nu_\gamma(T_{\gamma,h-1})} = \zeta\left(t^*_{\gamma,h-1}\right)$, $Z_{\nu_\gamma(T_{\gamma,h})} = \zeta\left(t^*_{\gamma,h}\right)$, we have independence between strata:

$$
\begin{aligned}
&h \neq h' \\
&\Rightarrow \mathrm{P}^{(S_{\gamma,h},S_{\gamma h'})|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right),Z_{\nu_\gamma(T_{\gamma h'-1})}=\zeta\left(t^*_{\gamma h'-1}\right),Z_{\nu_\gamma(T_{\gamma h'})}=\zeta\left(t^*_{\gamma h'}\right)} \\
&\quad = \mathrm{P}^{(S_{\gamma,h})|Z_{\nu_\gamma(T_{\gamma,h-1})}=\zeta\left(t^*_{\gamma,h-1}\right),Z_{\nu_\gamma(T_{\gamma,h})}=\zeta\left(t^*_{\gamma,h}\right)} \otimes P^{(S_{\gamma h'})|Z_{\nu_\gamma(T_{\gamma h'-1})}=\zeta\left(t^*_{\gamma h'-1}\right),Z_{\nu_\gamma(T_{\gamma h'})}=\zeta\left(t^*_{\gamma h'}\right)} \\
&\Rightarrow S_{\gamma,h} \text{ and } S_{\gamma h'} \text{ are independent conditionally on } T_{\gamma,h}, T_{\gamma,h-1}, T_{\gamma,h'} T_{\gamma,h'-1}.
\end{aligned}
$$

Together with equation (D.17), this implies that $\mathrm{P}^{\left(T_{\gamma h}\right)_{\gamma\in\mathbb{N},h\in\{1,\ldots,H-1\}}} - a.s \left(\left(t^*_{\gamma h}\right)_{\gamma\in\mathbb{N},h\in\{1,\ldots,H-1\}}\right)$,

$$
\mathrm{P}^{\sqrt{n_\gamma}\left(n_\gamma^{-1}S_\gamma - E_\infty\right)|T_{\gamma,1}=t^*_{\gamma,1},\ldots T_{\gamma,H-1}=t^*_{\gamma,H-1}} \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathcal{N}(0, V_\infty). \tag{D.27}
$$

The almost sure asymptotic normality implies the global asymptotic normality. For $\gamma \in \mathbb{N}$, $x \in \mathbb{R}$, $\alpha \in \mathbb{R}^p$ we define:

$$
h_{\gamma,\alpha,x} : t^* \mapsto \mathrm{E}\left[\exp\left(\mathrm{i}x\left(\sqrt{n_\gamma}\left(\frac{\alpha^{\mathrm{T}}\left(S_\gamma - E_\infty\right)}{n_\gamma\sqrt{\alpha^{\mathrm{T}} V_\infty \alpha}}\right)\right)\right)\middle| T_{\gamma,1} = t^*_{\gamma,1}, \ldots, T_{\gamma,H-1} = t^*_{\gamma,H-1}\right].
$$

Then equation (D.27) implies that $\mathrm{P}^{\left(T_{\gamma h}\right)_{\gamma\in\mathbb{N},h\in\{1,\ldots,H-1\}}} - a.s\,(t^*)$, $\lim_{\gamma\to\infty} h_{\gamma,\alpha,x,}(t^*) = \exp\left(\mathrm{i}x - t^2/2\right)$ Besides, $\mathrm{P}^{\left(T_{\gamma h}\right)_{\gamma\in\mathbb{N},h\in\{1,\ldots,H-1\}}} - a.s\,(t^*)$, $\forall\gamma \in \mathbb{N}\ |h_{\gamma,\alpha,x}(t^*)| \leqslant 1$. We apply the Lebesgue dominated convergence theorem:

$$
\begin{aligned}
&\lim_{\gamma\to\infty} \mathrm{E}\left[\exp\left(\mathrm{i}x\left(\sqrt{n_\gamma}\left(\left(n_\gamma\sqrt{\alpha^{\mathrm{T}} V_\infty \alpha}\right)^{-1}\alpha^{\mathrm{T}}\left(S_\gamma - E_\infty\right)\right)\right)\right)\right] \\
&= \lim_{\gamma\to\infty} \mathrm{E}\left[\mathrm{E}\left[\exp\left(\mathrm{i}x\left(\sqrt{n_\gamma}\left(\left(n_\gamma\sqrt{\alpha^{\mathrm{T}} V_\infty \alpha}\right)^{-1}\alpha^{\mathrm{T}}\left(S_\gamma - E_\infty\right)\right)\right)\right)\middle| T_{\gamma,1}=t^*_{\gamma,1},\ldots,T_{\gamma,H-1}=t^*_{\gamma,H-1}\right]\right] \\
&= \mathrm{E}\left[\lim_{\gamma\to\infty} \mathrm{E}\left[\exp\left(\mathrm{i}x\left(\sqrt{n_\gamma}\left(\left(n_\gamma\sqrt{\alpha^{\mathrm{T}} V_\infty \alpha}\right)^{-1}\alpha^{\mathrm{T}}\left(S_\gamma - E_\infty\right)\right)\right)\right)\middle| T_{\gamma,1}=t^*_{\gamma,1},\ldots,T_{\gamma,H-1}=t^*_{\gamma,H-1}\right]\right] \\
&= \int \exp\left(\mathrm{i}x - x^2/2\right) \mathrm{P}^{\left(T_{\gamma h}\right)_{\gamma\in\mathbb{N},h\in\{1,\ldots,H-1\}}}(t^*) \\
&= \exp\left(\mathrm{i}x - x^2/2\right).
\end{aligned}
$$

The convergence of the characteristic function implies the convergence to the normal distribution, which ends the demonstration of the theorem.

□

D.4.4 Proof of Result 5.2

We apply Lemma D.2 with

$$
g(y, z, \pi) = \begin{pmatrix} y \times z/\pi \\ y^2/\pi \end{pmatrix}.
$$

Then we obtain the asymptotic normality of the vector $\sqrt{n_\gamma} S_\gamma$:

$$
\sqrt{n_\gamma}\left(\begin{bmatrix} n_\gamma^{-1}\sum_{k=1}^{N_\gamma} Y_k Z_k/\pi_{\gamma k} I_{\gamma k} \\ n_\gamma^{-1}\sum_{k=1}^{N_\gamma} Y_k^2/\pi_{\gamma k} I_{\gamma k} \end{bmatrix} - E_\infty\right) \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathcal{N}(0, V_\infty),
$$

with $E_\infty = \left[\xi_0(\theta_0^2+1) \quad \theta_0^2+1\right]^{\mathrm{T}}$.

By applying the Delta method (see van der Vaart (1998, Theorem 3.1 p. 26)), we obtain that

$$\sqrt{n_\gamma}\left(\hat{\xi}_\gamma - \xi_0\right) \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}\left(\left((\theta_0^2+1)^{-1} \quad -\xi_0(\theta_0^2+1)^{-1}\right) V_\infty \begin{pmatrix} (\theta_0^2+1)^{-1} \\ -\xi_0(\theta_0^2+1)^{-1}\end{pmatrix}\right).$$

D.4.5 Proof of Result 5.4

We apply Lemma D.2 with:

$$g(y,z,\pi) = \begin{bmatrix} (\partial\Delta/\partial\theta)\,(y,\theta_0,\xi_0) \\ y\times z/\pi \\ y^2/\pi \end{bmatrix}.$$

Then we obtain the asymptotic normality of the vector $\sqrt{n_\gamma}\left(n_\gamma^{-1}S_\gamma - E_\infty\right)$:

$$\sqrt{n_\gamma}\begin{bmatrix} n_\gamma^{-1}\sum_{k=1}^{N_\gamma}(\partial\Delta/\partial\theta)\,(Y_k,\theta_0,\xi_0)I_{\gamma k} \\ n_\gamma^{-1}\sum_{k=1}^{N_\gamma} Y_k Z_k/\pi_{\gamma k} I_{\gamma k} \\ n_\gamma^{-1}\sum_{k=1}^{N_\gamma} Y_k^2/\pi_{\gamma k} I_{\gamma k}\end{bmatrix} \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}(E_\infty, V_\infty),$$

with $E_\infty = \left[0 \quad \xi(\theta^2+1) \quad \theta^2+1\right]^{\mathrm{T}}$. By applying the Delta method (see van der Vaart (1998, Theorem 3.1 p. 26)), we obtain that

$$\sqrt{n_\gamma}\begin{bmatrix} \left(\frac{\partial}{\partial\theta}\overline{\mathscr{L}}\right)\left(\left(Y_{R_\gamma(k)}\right)_{k\in\{1,\ldots,n_\gamma\}},\theta_0,\xi_0\right) \\ \hat{\xi}_\gamma - \xi_0\end{bmatrix}$$
$$\xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}\left(0, \begin{bmatrix}1 & 0 & 0 \\ 0 & (\theta_0^2+1)^{-1} & -\xi_0(\theta_0^2+1)^{-1}\end{bmatrix} V_\infty \begin{bmatrix}1 & 0 \\ 0 & (\theta_0^2+1)^{-1} \\ 0 & -\xi_0(\theta_0^2+1)^{-1}\end{bmatrix}\right).$$

D.4.6 Proof of Result 5.5

Proof. We apply Lemma 3 with: $g(y,z,\pi) = y/\pi$. Then we obtain the asymptotic normality of the vector $\bar{\theta} = \tau\sqrt{n_\gamma}\left(n_\gamma^{-1}\sum_{k=1}^{N_\gamma} I_{\gamma k}Y_k/\pi_{\gamma k}\right)$:

$$\sqrt{n_\gamma}\tau\left(n_\gamma^{-1}\sum_{k=1}^{N_\gamma} I_{\gamma k}Y_k/\pi_{\gamma k} - \mathrm{E}\left[Y\right]\right) \xrightarrow[\gamma\to\infty]{\mathscr{L}} \mathscr{N}(0, V_{\bar{\theta}}),$$

with

$$\begin{aligned} V_{\bar{\theta}} &= \tau\sum_{h=1}^{H}\left(t_{\infty h} - t_{\infty h-1}\right)\tau_{\infty h}\,\mathrm{Var}\left[Y_1/\tau_{\infty h}\middle|\, Z_1 \in \left]\zeta(t_{\infty h-1}),\zeta(t_{\infty h})\right]\right] \\ &= \tau\sum_{h=1}^{H}\left(t_{\infty h} - t_{\infty h-1}\right)\tau_{\infty h}^{-1}\,\mathrm{Var}\left[Y_1\middle|\, Z_1 \in \left]\zeta(t_{\infty h-1}),\zeta(t_{\infty h})\right]\right]. \end{aligned}$$

□

References

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Appendix E

Proofs for chapter 6

E.1 Proof of equation (6.14)

To simplify the notation, we note $b(\pi_k) \equiv b_k$. First note that the optimization problem is equivalent to find the vector $\mathbf{b} = (b_1, \ldots, b_N)'$ that minimizes

$$W_0(\mathbf{b}) = \sum_{k \in U} b_k \left(y_k - y_k^0(\mathbf{b})\right)^2$$

where $y_k^0(\mathbf{b}) = \mathbf{x}_k' \left(\sum_{l \in U} b_l \mathbf{x}_l \mathbf{x}_l'\right)^{-1} \sum_{l \in U} b_l \mathbf{x}_l y_l$, under the constraints:

$$b_k \geqslant 0 \text{ for any unit } k \in U \tag{E.1}$$

and

$$\sum_{k \in U} \frac{1}{b_k + 1} = n. \tag{E.2}$$

The partial derivative of $W_0(\mathbf{b})$, with respect to b_l, is equal to

$$\frac{\partial W_0(\mathbf{b})}{\partial b_l} = \left(y_l - y_l^0(\mathbf{b})\right)^2 + 2 \sum_{k \in U} b_k \left(y_k - y_k^0(\mathbf{b})\right) \frac{\partial \left(y_k - y_k^0(\mathbf{b})\right)}{\partial b_l}. \tag{E.3}$$

Since $y_k^0(\mathbf{b})$ may alternatively be written as $y_k^0(\mathbf{b}) = \mathbf{x}_k' \mathbf{A}(\mathbf{b})^{-1} \mathbf{c}(\mathbf{b})$, with $\mathbf{A}(\mathbf{b}) = \sum_{l \in U} b_l \mathbf{x}_l \mathbf{x}_l'$ and $\mathbf{c}(\mathbf{b}) = \sum_{l \in U} b_l \mathbf{x}_l y_l$, and since $\mathbf{x}_k$ does not depend on b_l, we have

$$\begin{aligned}
\tfrac{\partial y_k^0(\mathbf{b})}{\partial b_l} &= \mathbf{x}_k' \left(\tfrac{\partial (\mathbf{A}(\mathbf{b})^{-1})}{\partial b_l} \mathbf{c}(\mathbf{b}) + \mathbf{A}(\mathbf{b})^{-1} \tfrac{\partial \mathbf{c}(\mathbf{b})}{\partial b_l} \right) \\
&= \mathbf{x}_k' \left(-\mathbf{A}(\mathbf{b})^{-1} \tfrac{\partial \mathbf{A}(\mathbf{b})}{\partial b_l} \mathbf{A}(\mathbf{b})^{-1} \mathbf{c}(\mathbf{b}) + \mathbf{A}(\mathbf{b})^{-1} \tfrac{\partial \mathbf{c}(\mathbf{b})}{\partial b_l} \right) \\
&= \mathbf{x}_k' \mathbf{A}(\mathbf{b})^{-1} \mathbf{x}_l \left(y_l - \mathbf{x}_l' \mathbf{A}(\mathbf{b})^{-1} \mathbf{c}(\mathbf{b})\right).
\end{aligned}$$

By inserting this last expression into (E.3), we obtain

$$\begin{aligned}
\tfrac{\partial W_0(\mathbf{b})}{\partial b_l} &= \left(y_l - y_l^0(\mathbf{b})\right)^2 \\
&\quad -2 \left[\textstyle\sum_{k \in U} b_k (y_k - y_k^0(\mathbf{b})) \mathbf{x}_k'\right] \mathbf{A}(\mathbf{b})^{-1} \mathbf{x}_l \left(y_l - \mathbf{x}_l' \mathbf{A}(\mathbf{b})^{-1} \mathbf{c}(\mathbf{b})\right) \\
&= \left(y_l - y_l^0(\mathbf{b})\right)^2
\end{aligned}$$

since $\sum_{k \in U} b_k (y_k - y_k(\mathbf{b})) \mathbf{x}_k' = 0$. Then under the constraint (E.2), we get

$$\begin{aligned}
& \left(y_l - y_l^0(\mathbf{b})\right)^2 - \gamma \tfrac{1}{(b_l+1)^2} &&= 0 \\
\Leftrightarrow\ & \left(y_l - y_l^0(\mathbf{b})\right)^2 - \gamma \pi_l^2 &&= 0 \\
\Leftrightarrow\ & \pi_l &&= \sqrt{\gamma} |y_l - y_l^*(\pi)|
\end{aligned}$$

where γ denotes a Lagrange multiplier. The result follows by application of constraint (6.13).

E.2 Proof of Theorem 6.1

For any $t = 0, \ldots, T$, denote $\mathbf{b}^t = \left[b(\alpha_1^t), \ldots, b(\alpha_i^t), \ldots, b(\alpha_I^t)\right]^{\mathrm{T}}$. Let $\mathbf{u} = \left[u_1, \ldots, u_i, \ldots, u_I\right]^{\mathrm{T}}$ be any $I \times 1$ vector, and

$$W_1(\mathbf{u}) = \frac{N}{N-q} \sum_{i=1}^{I} N_i u_i \sigma_i^2(\boldsymbol{\alpha}^{t-1}).$$

The minimization of $W_1(\mathbf{u})$ in $\mathbf{u}$, subject to

$$\sum_{i=1}^{I} \frac{N_i}{u_i + 1} = n \tag{E.4}$$

leads to $\mathbf{u} = \mathbf{b}^t$. Since $\mathbf{b}^{t-1}$ also satisfies equation (E.4), we have

$$W_1(\mathbf{b}^t) \leqslant W_1(\mathbf{b}^{t-1}) = V(\boldsymbol{\alpha}^{t-1}). \tag{E.5}$$

Now, let

$$W_2(\boldsymbol{\beta}) = \frac{N}{N-q} \sum_{i=1}^{I} b(\alpha_i^t) \sum_{k \in U_i} (y_k - \mathbf{x}_k^{\mathrm{T}} \boldsymbol{\beta})^2$$

where $\boldsymbol{\beta}$ denotes a $q \times 1$ vector. This is a standard fact that $W_2(\boldsymbol{\beta})$ is minimized by $\boldsymbol{\beta}^t = \left(\sum_{i=1}^{I} b(\alpha_i^t)\mathbf{A}_i\right)^{-1} \sum_{i=1}^{I} b(\alpha_i^t)\mathbf{c}_{1i}(y)$. Consequently, we obtain

$$W_2(\boldsymbol{\beta}^t) \leqslant W_2(\boldsymbol{\beta}^{t-1}) \tag{E.6}$$

where $\boldsymbol{\beta}^{t-1} = \left(\sum_{i=1}^{I} b(\alpha_i^{t-1})\mathbf{A}_i\right)^{-1} \sum_{i=1}^{I} b(\alpha_i^{t-1})\mathbf{c}_i(y)$. Since $W_2(\boldsymbol{\beta}^t) = V(\boldsymbol{\alpha}^t)$ and $W_2(\boldsymbol{\beta}^{t-1}) = W_1(\mathbf{b}^t)$, the result follows by a joint application of (E.5) and (E.6).

Appendix F

-code used for simulations

```
##0. Libraries
library(np)
library(ks)
# library(sampling)
require(tikzDevice)
# load(paste(getwd(),"/.RData")
setwd("~/Dropbox/Travail/Recherche/Travaux/MÃ1'moire de thÃÍse/v10/Figures")
# save.image()
##Chapitre 1.
##1.1. Definition of empirical cumulative distribution associated to a vector
FDR     <-function(y){return(function(x){return(lapply(x,function(x,y){round(
    sum(y<=x)/length(y),4)},y=y))})}
##1.2. Definition of a kernel density estimator applied to a vector
densite<-function(y){return(function(x){round(as.vector(kde(y,0.005,eval.
    points=x)$estimate),4)})}
##1.3. definition of some sampling designs
SRS<-function(tau){
  list(
    S=function(z){sample(1:length(z),size=floor(tau*length(z)),replace=F)},
    Pik=function(z){rep(floor(tau*length(z))/length(z),length(z))},
    demarc=function(z){NULL})}
StratS<-function(proph,tauh){
  list(
    Pik=function(z){
      oo<-rank(z);
      N=length(z)
      Nh<-vector()
      nh<-vector()
      Pikk<-rep(NULL,length(z));cum=0
      for (i in 1:(length(proph)-1)){
        Nh[i]<-floor(proph[i]*N);
        nh[i]<-floor(tauh[i]*Nh[i]);
        Pikk[oo>cum&oo<=cum+Nh[i]]<-nh[i]/Nh[i];cum<-cum+Nh[i]}
      Nh[i+1]<-N-sum(Nh);nh[i+1]<-floor(tauh[i+1]*Nh[i+1]);Pikk[oo>cum]<-nh[i
          +1]/Nh[i+1];
      return(Pikk)},
    S=function(z){
      oo<-rank(z);
```

```
      N=length(z)
      Nh<-vector()
      nh<-vector()
      S<-vector();cum<-0
      for (i in 1:(length(proph)-1)){Nh[i]<-floor(proph[i]*N);nh[i]<-floor(
          tauh[i]*Nh[i]);
      S=c(S,sample((1:N)[(oo>cum)&(oo<=cum+Nh[i])],size=nh[i],replace=FALSE));
          cum<-cum+Nh[i]}
      Nh[i+1]<-N-sum(Nh);nh[i+1]<-floor(tauh[i+1]*Nh[i+1]);S=c(S,sample((1:N)
          [(oo>cum)&(oo<=N)],size=nh[i+1],replace=FALSE))
      return(S)},
    param=list(proph=proph,tauh=tauh),
    demarc=function(z){
      z<-sort(z);
      N=length(z)
      dem<-vector()
      Nh<-vector();cum<-0;
      for (i in 1:(length(proph)-1)){Nh[i]<-floor(proph[i]*N);cum<-cum+Nh[i];
          dem[i]<-sort(z)[cum]}
      return(dem)})}
SWRPPS<-function(tau){
  list(
    Pik=function(z){z/sum(z)},
    S=function(z){sample(1:length(z),size=floor(tau*length(z)),prob=1-(1-z/sum
        (z))^(floor(tau*length(z))),replace=TRUE)},
    demarc=function(z){NULL})}
ClusterS<-function(proph,probh,tauh){
    list(
      Pik=function(z){
            oo<-order(z);
            N=length(z)
            Nh<-vector()
            nh<-vector()
            Pikk<-rep(NULL,length(z));cum=0
            for (i in 1:(length(proph)-1)){Nh[i]<-floor(proph[i]*N);
            cum<-cum+Nh[i];nh[i]<-floor(tauh[i]*Nh[i]);Pikk[oo>cum&oo<=cum+Nh[
                i]]<-proph[i]*(nh[i]/Nh[i]);}
            Nh[i+1]<-N-sum(Nh);nh[i+1]<-floor(tauh[i+1]*Nh[i+1]);Pikk[oo>cum]
                <-proph[i]*(nh[i+1]/Nh[i+1]);
            return(Pikk)},
      S=function(z){
            oo<-order(z);
            N=length(z)
            nbclus<-length(proph)
            clus<-sample(1:nbclus,size=1,prob=probh,replace=FALSE);
            cum<-floor((sum(proph[1:clus])-proph[clus])*N)
            Nh<-floor(proph[clus]*N);
            S=sample((1:N)[oo>=cum&oo<cum+Nh],size=floor(tauh[clus]*Nh),
                replace=F)
            return(S)},
        demarc=function(z){
            oo<-order(z);
            N=length(z)
```

```
                dem<-vector()
                Nh<-vector();cum<-0;
                for (i in 1:(length(proph)-1)){Nh[i]<-floor(proph[i]*N);cum<-cum+
                    Nh[i];dem[i]<-z[oo==cum]}
                return(dem)})}
##1.4.population and sampling design models
model.unif.pps<-function(tau){
  list(rloiy=function(N){2*runif(N)},
  ploi=function(y){punif(y/2)},
  ploilim=function(y){y/2*(y>0)*(y<=2)},
  rloiz=function(y){y},
  dloi=function(y){dunif(y/2)/2},
  Scheme=SWRPPS(tau),
  rho=function(y){y},
  vinf=function(y){tau*y},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.1,2.1))}
model.chisq.pps<-function(tau){
  list(rloiy=function(N){rchisq(N,1)},
  ploi=function(y){pchisq(y,1)},
  ploilim=function(y){pgamma(y,3/2,2)},
  rloiz=function(y){y},
  dloi=function(y){dchisq(y,1)},
  Scheme=SWRPPS(tau),
  rho=function(y){y},
  vinf=function(y){tau*y},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.1,4.1))}
model.chisq2.pps<-function(tau){
  list(rloiy=function(N){rchisq(N,1)},
  ploi=function(y){pchisq(y,1)},
  ploilim=function(y){(pgamma(y,3/2,2)+pgamma(y,5/2,2))/2},
  rloiz=function(y){y+rchisq(N,1)},
  dloi=function(y){dchisq(y,1)},
  Scheme=SWRPPS(tau),
  rho=function(y){(y+1)/2},
  vinf=function(y){tau*y},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.1,2.1))}
model.indep.strat<-function(proph,tauh){
  tau=sum(proph*tauh)
  return(list(
  param=list(proph=proph,tauh=tauh),
  rloiy=function(N){rnorm(N,mean=2,sd=1)},
  ploi=function(y){pnorm(y,mean=2,sd=1)},
  ploilim=function(y){pnorm(y,mean=2,sd=1)},
  rloiz=function(y){rnorm(length(y),mean=2,sd=1)},
  dloi=function(y){dnorm(y,mean=2,sd=1)},
  Scheme=StratS(proph,tauh),
  rho=function(y){1},
```

```
  vinf=function(y){tau-tau^2},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.5,5)))}
model.unif.SRS<-function(tau){
  return(list(
  rloiy=function(N){runif(N)},
  ploi=function(y){punif(y)},
  rloiz=function(y){rnorm(length(y))},
  dloi=function(y){dunif(y)},
  Scheme=SRS(tau),
  rho=function(y){1},
  vinf=function(y){tau-tau^2},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.1,1.1)))}
model.norm.cluster<-function(proph,probh,tauh,xi){
  tau=sum(proph*probh*tauh)
  Ninf=100000;Nhinf=floor(Ninf*proph);
  yinf<-rnorm(Ninf)+2;
  sinf<-StratS(proph*probh,tauh)$S(xi*yinf+rnorm(Ninf));
  dloilim<-densite(yinf[sinf])
  return(list(
  param=list(proph=proph,probh=probh,tauh=tauh,xi=xi),
  rloiy=function(N){rnorm(N,2)},
  ploi=function(y){pnorm(y,2)},
  rloiz=function(y){rnorm(y,mean=xi*y,sd=1)},
  dloi=function(y){dnorm(y,2)},
  ploilim=FDR(yinf[sinf]),
  dloilim=dloilim,
  Scheme=ClusterS(proph,tauh,probh),
  rho=function(y){dloilim(y)/dnorm(y,2)},
  vinf=function(y){NULL},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.5,4)))}
model.dep.strat<-function(proph,tauh,theta,xi,sigma){
  Ninf=100000;Nhinf=floor(Ninf*proph);
  yinf<-rnorm(Ninf)+2;
  sinf<-StratS(proph,tauh)$S(rnorm(yinf,mean=xi*yinf,sd=sigma));
  tau=sum(proph*tauh)
    rho=function(y){
        rhorho<-tauh[length(tauh)];t<-proph[1]
      for(h in 1:(length(tauh)-1)){rhorho<-rhorho+(tauh[h]-tauh[h+1])*pnorm((
        sqrt(xi^2+sigma^2)*qnorm(t)+xi*(theta-y))/sigma);t<-t+proph[h+1]}
      return(rhorho/tau)}
  return(list(
  param=list(proph=proph,tauh=tauh,theta=theta,xi=xi,sigma=sigma),
  rloiy=function(N){rnorm(N,theta,1)},
  ploi=function(y){pnorm(y,theta,1)},
  rloiz=function(y){rnorm(y,mean=xi*y,sd=sigma)},
  dloi=function(y){dnorm(y,theta,1)},
  ploilim=FDR(yinf[sinf]),
```

```
  dloilim=function(y){return(rho(y)*dnorm(y,theta,1))},
  Scheme=StratS(proph,tauh),
  rho=rho,
  vinf=function(y){tau*rho(y)-(tau*rho(y))^2},
  En=function(N){tau*N},
  tau=tau,
  supportY=c(-.5,5)))}
##1.5. Population and sample realisation
genere<-function(m=m,N=1000,Y=NA){
  Scheme=m$Scheme
  Yg<-m$rloiy(N);if(!is.na(Y)){Yg<-Y}
  Zg<-m$rloiz(Yg);
  pik<-Scheme$Pik(Zg);
  Sg<-Scheme$S(Zg);
  demarc<-Scheme$demarc(Zg);
  Ig<-rep(0,N);for(k in unique(Sg)){Ig[k]<-sum(Sg==k)}
  NHT=sum(1/pik[Sg])
  n=sum(Ig)
Gg=list(N=N,Yg=Yg,Zg=Zg,pik=pik,Sg=Sg,Ig=Ig,n=n,NHT=NHT,demarc=demarc)
return(Gg)
}
##1.6. Code tex pour graphiques
##1.6.1. fonctions gÃľnÃľriques de tracÃľ.
#ouverture de fichier
debutfic<-function(fic,echellegraph=c(1,1),limites=c(-0.35,-0.13,1.1,1.1)){
    write(paste("\\begin{tikzpicture}[line cap=round,line join=round,>=
       triangle 45,x=",echellegraph[1],"cm,y=",echellegraph[2],"cm]"),file=
       fic,append=F)
    write(paste("\\clip(",limites[1],",",limites[2],") rectangle (",limites
       [3],",",limites[4],");"),file=fic,append=T)}
#tracÃľ des axes
traceaxes<-function(fic,limites=c(-0.35,-0.13,1.1,1.1),width=1.2){
    write("%%Les axes",file=fic,append=T)
    styleaxe="->";aj="";if (styleaxe=="->"){aj<-"->,"}
    write(paste("\\draw[",aj,"line width=",width,"pt,color=black] (",limites
       [1],",0) --  (",limites[3],",0);"),file=fic,append=T)
    write(paste("\\draw[",aj,"line width=",width,"pt,color=black] (0,",limites
       [2],") -- (0,",limites[4],");"),file=fic,append=T)}
#tracÃľ de courbes
tracecourbe<-function(fic,x,Y,couleurcourbe=c("black"),width=1.5){
    write("%Les courbes",file=fic,append=T)
    for(i in 1:length(Y[1,]))){xc<-x[!is.na(x)&!is.na(Y[,i])];yc<-(Y[,i])[!is.
       na(x)&!is.na(Y[,i])];
    if (sum(!is.na(yc))>2){lb<-(max(length(yc),length(xc))-1)
    write(paste("%%courbe ",i),file=fic,append=T)
    for (j in 1:lb){
      write(paste("\\draw[line width=",width,"pt,color=",couleurcourbe[i],
                  "] (",  formatC(xc[j],    format="f"),",",formatC(yc[j
                      ],  format="f"),
                  ")-- (",formatC(xc[j+1],format="f"),",",formatC(yc[j
                      +1],format="f"),");"),file=fic,append=T)}}}
}
#tracÃľ d'une fonction de rÃľpartition empirique
```

```
tracefdr<-function(fic,couleur="black",y,limites,width=2){
    write("%FDR",file=fic,append=T)
    nn=length(y)
    ys=c(limites[1],sort(y),limites[2]);
    for(i in (1:(nn+1))){
      if(ys[i]!=ys[i+1]){
    if (i < nn+1){
    #  write(paste("\\draw [color=",couleur,"] (",ys[i+1],",",(i-1)/nn,")
       circle (3pt);"),file=fic,append=T)
        }}
    if (i >1){
     # if (ys[i]<ys[i+1]){write(paste("\\fill [color=",couleur,"] (",ys[i
       ],",",(i-1)/nn,") circle (3pt);"),file=fic,append=T)}
    }}
    for(i in (1:(nn+1))){
      if(ys[i]!=ys[i+1]){
    write(paste("\\draw [line width=",width,"pt,color=",couleur,"] (",formatC(
       ys[i],format="f"),",",formatC((i-1)/nn,format="f"),") -- (",formatC(ys
       [i+1],format="f"),",",formatC((i-1)/nn,format="f"),");"),file=fic,
       append=T)}}}
#tracÃ© de graduations sur les axes
tracegrades<-function(fic,repx="",repy=""){
    write("%grades",file=fic,append=T)
      xp=paste(repx,collapse=",");
      if(xp!=""){
    write(paste("\\foreach \\x in {",xp,"}"),file=fic,append=T)
    write(paste("\\draw[shift={(\\x,0)},color=black] (0pt,2pt) -- (0pt,-2pt)
       node[below] {\\footnotesize $\\x$};"),file=fic,append=T)}
      yp=paste(repy,collapse=",");
      if(yp!=""){
    write(paste("\\foreach \\y in {",yp,"}"),file=fic,append=T)
    write(paste("\\draw[shift={(0,\\y)},color=black] (2pt,0pt) -- (-2pt,0pt)
       node[left] {\\footnotesize $\\y$};"),file=fic,append=T)}}
# Ã©criture de texte sur le graphique
tracetexte<-function(fic,texte=c("$z_{\\gamma k}$","$Y_k$"),posy=posy,posz=
    posz){
    write("%texte",file=fic,append=T)
    for (i in 1:length(texte)){write(paste("\\draw[color=black] (",posy[i],","
       ,posz[i],") node {",texte[i],"};"),file=fic,append=T)}}
# tracÃ© de points
tracepts<-function(fic,y,z,color="black",diam=1){
    write("%points",file=fic,append=T)
      for(i in 1:length(y)){write(paste("\\fill[color=",color,"] (",formatC(y[
         i],format="f"),",",formatC(z[i],format="f"),") circle (",diam,"pt);"
         ),file=fic,append=T)}}
# tracÃ© de points
traceptsrep<-function(fic,y,z,s,couleurpoints="gray",diam=3){
    write("%points selectionnÃ©s",file=fic,append=T)
      for(k in unique(s)){
    ik<-sum(s==k)
    if(ik>1){
    write(paste("\\fill[color=",couleurpoints,"] (",formatC(y[k],format="f"),"
       ,",formatC(z[k],format="f"),") circle (",diam,"pt) node[below] {\\
```

```
        footnotesize $",sum(s==k),"$};"),file=fic,append=T)}
    if(ik==1){
    write(paste("\\fill[color=",couleurpoints,"] (",formatC(y[k],format="f"),"
        ,",formatC(z[k],format="f"),") circle (",diam,"pt);"),file=fic,append=
        T)}
    }}
# tracÃ© de lignes en pointillÃ©s
tracepointillesh<-function(fic,demarc,color="gray",limites,width=2.4){
  if (!is.null(demarc)){
    write("%lignes",file=fic,append=T)
    for(demarq in demarc){
    write(paste("\\draw[line width=",width,"pt,dash pattern=on 2pt off 2pt,
        color=gray] (0,",demarq,") -- (",limites[3],",",demarq,");%axe de
        separation des strates"),file=fic,append=T)}}}#separation des strates
tracepointillesv<-function(fic,demarc,color="gray",limites,width=2.4){
  if (!is.null(demarc)){
    write("%lignes",file=fic,append=T)
    write(paste("\\draw[line width=",width,"pt,dash pattern=on 2pt off 2pt,
        color=gray] (",demarc,",0) -- (",demarc,",",limites[4],");%axe de
        separation des strates"),file=fic,append=T)}}#separation des strates
# gÃ©nÃ©ration de code pour un graphique
generecodetexgraph<-function(fic,y,s,z,taillegraph,prop,quoi,x,Y,repx=repx,
    repy=repy,leg,couleurs=couleurs,demarc=NULL,tailles=tailles,maxy,maxz){
        limites=prop*c(maxy,maxz,maxy,maxz)
        echellegraph<-taillegraph/c(limites[3]-limites[1],limites[4]-limites
            [2])
        leg$posy<-leg$posy*maxy
        leg$posz<-leg$posz*maxz
        if (max(quoi=="deb")){debutfic(fic,echellegraph=echellegraph,limites=
            limites)}
        if (max(quoi=="axes")){traceaxes(fic,limites=limites,width=tailles$axe
            )}
        if (max(quoi=="courbe")){tracecourbe(fic,x,Y,couleurcourbe=couleurs$
            courbe,width=tailles$courbe)}
        if (max(quoi=="fdr")){tracefdr(fic,couleur=couleurs$cdf,y,limites[c
            (1,3)],width=tailles$cdf)}
        if (max(quoi=="fdremp")){tracefdr(fic,couleur=couleurs$scdf,y[s],
            limites[c(1,3)],width=tailles$scdf)}
        if (max(quoi=="grades")){tracegrades(fic,repx=repx,repy=repy)}
        if (max(quoi=="ppts")){tracepts(fic,y,z,color=couleurs$pts,diam=
            tailles$pts)}
        if (max(quoi=="spts")){traceptsrep(fic,y,z,s,couleurpoints=couleurs$
            ptss,diam=tailles$ptss)}
        if (max(quoi=="texte")){tracetexte(fic,texte=leg$texte,posy=leg$posy,
            posz=leg$posz)}
        if (max(quoi=="lignev")){tracepointillesv(fic,demarc,color=couleurs$
            ptl,limites,width=tailles$ptl)}
        if (max(quoi=="ligneh")){tracepointillesh(fic,demarc,color=couleurs$
            ptl,limites,width=tailles$ptl)}
        if (max(quoi=="strate")){
            write("\\usefont{T1}{ptm}{m}{n}",file=fic,append=T);
            write("\\rput{-270.0}(-0.05,0.2){\\rput(0.1,0){strata 1}};",file
                =fic,append=T)
```

```
              write("\\usefont{T1}{ptm}{m}{n}",file=fic,append=T);
              write("\\rput{-270.0}(-0.05,0.6){\\rput(0.1,0){strata 2}};",file
                 =fic,append=T)}
         if (max(quoi=="fin")){write("\\end{tikzpicture}",file=fic,append=T)}}
##1.6.2 Graphic generation function
#default parameters
legende<-list(
    list(posy=c(0.09,0.95),posz=c(1,0.1),texte=c("$Z_{k}$","$Y_k$"),col=rep("
       black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(0.12,0.45),posz=c(0.8,0.55,0.8,0.55,.1),texte=c("$F_{\\infty}$
       ","$Y_k$"),col=rep("black",2)),
    list(posy=c(1),posz=c(.1),texte="",col="black"))
  taillegraph<-list(c(4,5),c(8,5),c(8,5),c(8,5),c(8,5),c(8,5),c(8,5),c(8,5))
  listcouleur<-list(sample="orangeensai",pop="black",lim="bleuensai1",soft="
     gray")
  listtaille<-list(fin=0.6,axes=0.6,epais=1.2,ptss=2,pts=.6)
  repy=list(NULL,NULL,NULL,NULL,NULL,NULL,NULL,NULL,NULL);
  repz<-repy
#procedure
plotgen<-function(fic3,Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,
    listcouleur=listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,maxz3,
    quoigen){
  tailles<-list(ptl=listtaille$fin,axes=listtaille$axes,courbe=rep(listtaille$
     fin,3),scdf=listtaille$epais,cdf=listtaille$epais,ptss=listtaille$ptss,
     pts=listtaille$pts)
  couleurs<-list(ptss=listcouleur$sample,pts=listcouleur$pop,cdf=listcouleur$
     pop,scdf=listcouleur$sample,lim=listcouleur$lim,ptl=listcouleur$soft,
     courbe=c(listcouleur$pop,listcouleur$sample,listcouleur$lim))

  x<-round(maxy*seq(prop1[1],prop2[3],length.out = 300),4);
  #a.
  if(max(quoigen==1)|max(quoigen==2)|max(quoigen==3)|max(quoigen==4)){Y<-cbind
     (m$ploi(x),m$ploilim(x),m$ploilim(x))}
  if(max(quoigen==1)){fic=paste(fic3,"a.tex",sep="");
  quoi<-c("deb","axes","spts","texte","grades","ppts","strate","ligneh","fin")
     ;
  generecodetexgraph(fic,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[1]],prop2,quoi,x,Y,
     repy[[1]],repz[[1]],legende[[1]],couleurs,Gg$demarc,tailles,maxy,maxz)}
  #b.limit sample cdf and population cdf
   if(max(quoigen==2)){ quoi=c("deb","axes","courbe","texte","grades","fin");
  fic=paste(fic3,"b.tex",sep="");
  texte<-c("$F$","$F_\\infty$"); tailles$courbe<-rep(listtaille$epais,2)
```

```
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[2]],prop1,quoi,x,Y,
      repy[[2]],repz[[2]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)}
  #c.empirical population cdf
  if(max(quoigen==3)){quoi<-c("deb","axes","courbe","fdr","grades","fin");
  fic<-paste(fic3,"c.tex",sep="");  tailles$courbe<-rep(listtaille$fin,2)
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[3]],prop1,quoi,x,Y,
      repy[[3]],repz[[3]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)}
  #d.empirical sample cdf
  if(max(quoigen==4)){quoi=c("deb","axes","courbe","texte","fdremp","grades","
      fin")
  fic=paste(fic3,"d.tex",sep="");
  texte<-c("$\\alpha$");texte<-c("")
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[4]],prop1,quoi,x,Y,
      repy[[4]],repz[[4]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)}

  if(max(quoigen==5)){kdey<-function(x){round(as.vector(kde(Gg$Yg[Gg$Sg],hpi(x
      =Gg$Yg[Gg$Sg])/2,eval.points=x)$estimate),4)}
  Y<-cbind(m$dloi(x),kdey(x),m$dloi(x)*m$rho(x));
  quoi=c("deb","axes","courbe","texte","fin")
  fic=paste(fic3,"e.tex",sep="");
  echelled<-max(Y)
  texte<-c("$\\alpha$");texte<-c("")
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[5]],prop1,quoi,x,Y,
      repy[[5]],repz[[5]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)
}

  y0<-x
  #Gg<-list(Yg=y,Sg=s,N=length(y),n=length(y[s]));ker=kergaus;m
  if(max(quoigen==6)){Y<-cbind(m$dloi(x),p2sr(y0,Gg,b,ker=ker,m=m),fHT(y0,Gg,b
      ,ker=ker,m=m));
  quoi=c("deb","axes","courbe","texte","fin")
  fic=paste(fic3,"f.tex",sep="");
  texte<-c("$\\alpha$");
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[6]],prop1,quoi,x,Y,
      repy[[6]],repz[[6]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)}
  if(max(quoigen==7)){fic=paste(fic3,"g.tex",sep="");
  Y<-cbind(m$dloi(x),p2sr(y0,Gg,b,ker=ker,m=m),p2(y0,Gg,b,ker=ker,m=m));
  generecodetexgraph(fic ,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[7]],prop1,quoi,x,Y,
      repy[[7]],repz[[7]],legende[[2]],couleurs,Gg$demarc,tailles,maxy,1)}
  if(max(quoigen==8)){
  couleurs$courbe<-c(listcouleur$lim,listcouleur$sample)
    quoi=c("deb","axes","courbe","texte","fin")
    y0<-seq(m$support[1],m$support[2],length.out=100);
    mmts<-moments(y0,b,ker=ker,m=m,lafun=p,N=Gg$N,nrep=1000);
    vpemp<-mmts$var;
    vp<-varp(y0,b=b,ker=ker,m=m,N=N);
    maxx=max(c(vpemp,vp))
    fic<-paste(fic3,"h.tex",sep="");
    leg<-list(posy=c(0),posz=c(1),texte=c(paste(maxx,"",sep="")),col=c("black"
        ))
    Y<-cbind(vpemp,vp)/maxx
    repz=c(1)
```

```
    generecodetexgraph(fic,Gg$Yg,Gg$Sg,Gg$Zg,taillegraph[[8]],prop1,quoi,y0,Y,
        repy,repz,leg,couleurs,Gg$demarc,tailles,maxy,1)
  }}



##Chapitre 4. Kernel density estimation
##4.1. Definitions
# Kernels
kergaus<-list(K=dnorm,intK2=(1/(2*sqrt(pi))));
ker<-kergaus

# bandwidths
b<-function(N){1/sqrt(N)}
varp<-function(y0,b,ker=ker,m=m,N){(m$dloi(y0)/(N*b(N)))*(m$vinf(y0)/(m$tau^2)
    +((m$rho(y0))^2))*ker$intK2}
varpsr<-function(y0,b,ker=ker,m=m,N){varp(y0,b,ker=ker,m=m,N)/((m$rho(y0))^2)}

# Kernel density estimators definition of kde
p0  <-function(y0,Gg,b,ker=ker,m=m){sum(ker$K((Gg$Yg[Gg$Sg]-y0)/b(Gg$N)))/(b(
    Gg$N)*Gg$n)}
fHT0<-function(y0,Gg,b,ker=ker,m=m){sum(ker$K((Gg$Yg[Gg$Sg]-y0)/hpi(x=Gg$Yg[Gg
    $Sg])*2)/Gg$pik[Gg$Sg])/(b(Gg$N)*Gg$NHT)*2}
p   <-function(y0,Gg,b,ker=ker,m=m){sapply(y0,p0,Gg=Gg,b=b,ker=ker,m=m)}
p2  <-function(y0,Gg,b,ker=ker,m=m){as.vector(kde(Gg$Yg[Gg$Sg],hpi(x=Gg$Yg[Gg$
    Sg])/2,eval.points=y0)$estimate)}
psr <-function(y0,Gg,b,ker=ker,m=m){p(y0,Gg,b,ker=ker,m=m)/m$rho(y0)}
p2sr<-function(y0,Gg,b,ker=ker,m=m){p2(y0,Gg,b,ker=ker,m=m)/m$rho(y0)}
fHT <-function(y0,Gg,b,ker=ker,m=m){sapply(y0,fHT0,Gg=Gg,b=b,ker=ker,m=m)}

# Calculus of variance
moments<-function(y0,b,ker=ker,m=m,lafun=psr,N=N,nrep=1000){
  XX<-matrix(NA,nrep,length(y0))
  for(i in 1:nrep){
  Gg<-genere(m,N)
  XX[i,]<-lafun(y0,Gg,b,ker=ker,m=m)}
  return(list(E=apply(XX,2,mean),var=apply(XX,2,var)))}
##Simulations and verification of variance formula
Verif<-function(m,N,b,nrep=100,nbpts=30,fic,verifvp){
  y0<-seq(m$support[1],m$support[2],length.out=nbpts);
  mmts<-moments(y0,b,ker=ker,m=m,lafun=p,N=N,nrep=nrep);
  vpemp<-mmts$var;
  vp<-varp(y0,b=b,ker=ker,m=m,N=N);
  png(paste(fic,"_1.png"))
    plot(y0,vpemp/vp,type='l',col="blue");
    title("v empirique/v theorique")
  dev.off()
  png(paste(fic,"_2.png"))
    plot(y0,vp,type='l',col="blue");
    points(y0,vpemp,type='l',col="orange");
    points(y0,verifvp(y0),type='l',col="red");
    title("v empirique(orange) - v theorique (bleu) - v theorique verif (rouge
        )")
```

```
  dev.off()
  png(paste(fic,"_3.png"))
    plot(y0,m$dloi(y0),type='l',col="black");
    plot(y0,m$rho(y0)*m$dloi(y0),type='l',col="blue");
    points(y0,mmts$E,type='l',col="orange");
    title("E[p] empirique(orange) -  f (bleu)- rho f (noir)")
  dev.off()
  png(paste(fic,"_4.png"))
    Gg<-genere(m,N)
    plot(y0,m$rho(y0)*m$dloi(y0),type='l',col="black");
    points(y0,p(y0,Gg,b,ker=ker,m=m),type='l',col="orange");
    title("p orange -  rho f (bleu)- rho f (noir)")
    dev.off()
return(list(vp=vp,vpemp=vpemp,y0=y0,mmts=mmts,m=m,verifvp=verifvp,N=N,nrep=
    nrep))
}

##4.2.Examples
#general parameters for graphics
nrep=10;nbpts=100;
##4.2.1. Independent sampling
#parameters
nrep=1000;nbpts=20;N<-10000;b<-function(N){0.2};ker<-kergaus;
proph<-c(.7,.3);tauh<-c(.5,.8);m<-model.indep.strat(proph,tauh)
    y0<-seq(m$support[1],m$support[2],length.out=100);
    mmts<-moments(y0,b,ker=ker,m=m,lafun=p,N=Gg$N,nrep=1000);
    vpemp<-mmts$var;
    vp<-varp(y0,b=b,ker=ker,m=m,N=N);
    maxx=max(c(vpemp,vp))
tikz('ish.tex')
plot(y0,vpemp,col="black",type='l',xlab="",ylab="");points(y0,vp,col="gray",
    type='l')
axis(1, at=x,labels=x,  las=0,pos=0)
axis(2, at=x,labels=x,  las=2,pos=0)
dev.off()
##4.2.2. Cluster sampling

##4.2.3. With replacement sampling
#global parameters
tau=.125; m<-model.chisq2.pps(tau);b<-function(N){0.1};ker<-kergaus;
#graphic1
N<-200;dodo=TRUE;k=1;while(dodo){k=k+1;Gg<-genere(m=m,N=N);dodo<-(max(Gg$Ig)
    ==1)&(k<5000)}
    maxy=0.85*max(Gg$Yg);maxz=0.85*max(Gg$Zg);
prop1<-c(-0.1,-0.12,1.05,1.2);prop2<-c(-0.1,-0.12,1.05,1.2);
plotgen("wr",Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,listcouleur=
    listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,maxz3,quoigen=c(1,4)
    )
N<-10000;Gg<-genere(m=m,N=N);    maxy=0.8*max(Gg$Yg);maxz=0.85*max(Gg$Zg);
plotgen("wr",Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,listcouleur=
    listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,maxz3,quoigen=c
    (2,3,5,6,7))
#graphic2
```

```
tau<-0.2
N<-10000;
 y0<-seq(m$support[1],m$support[2],length.out=nrep);
     mmts<-moments(y0,b,ker=ker,m=m,lafun=p,N=N,nrep=1000);
     vpemp<-mmts$var;
     vp<-varp(y0,b=b,ker=ker,m=m,N=N);
tikz('wrh.tex')
plot(y0,vpemp,col="black",type='l',xlab="",ylab="");points(y0,vp,col="gray",
    type='l')
axis(1, at=x,labels=x,  las=0,pos=0)
axis(2, at=x,labels=x,  las=2,pos=0)
dev.off()

#Verification of formula.
# tau=.3;m=model.unif.SRS(tau);N=5000;nrep=15000;nbpts=50;fic="test1";verifvp=
    function(y){((y>0)*(y<1)/(tau*N*b(N)))/(2*sqrt(pi))};
# sim1<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
#
# tau<-.65;m<-model.indep.strat(c(.5,.5),c(.5,.8));N=5000;nrep=30000;nbpts=50;
    fic="test2";verifvp=function(y){(dnorm(y,2)/(tau*N*b(N)))/(2*sqrt(pi))};
# sim2<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
#
# tau<-.5;m<-model.unif.pps(tau);N=5000;nrep=30000;nbpts=50;fic="test3";
    verifvp=function(y){(((y>0)*(y<2)/2)/(N*b(N))*(y/tau +y^2))/(2*sqrt(pi))};
# sim3<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
#
#
# tau=.3;m=model.unif.SRS(tau);N=10000;nrep=30000;nbpts=100;fic="test4";
    verifvp=function(y){((y>0)*(y<1)/(tau*N*b(N)))/(2*sqrt(pi))};
# sim4<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
# save(sim4,file="sim4")

tau<-.65;b<-function(N){.15};m<-model.indep.strat(c(.5,.5),c(.5,.8));N=25000;
    nrep=3;nbpts=100;fic="test5";
verifvp=function(y){(dnorm(y,2)/(tau*N*b(N)))/(2*sqrt(pi))};
sim5<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
save(sim5,file="sim5")
# .05, ... ,.5
tau<-.5;b<-function(N){.15};m<-model.unif.pps(tau);N=50000;nrep=3;nbpts=100;
    fic="test6";verifvp=function(y){(((y>0)*(y<2)/2)/(N*b(N))*(y/tau +y^2))/(2
    *sqrt(pi))};
sim6<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
save(sim6,file="sim6")


tau<-.5;b<-function(N){.05};m<-model.chisq2.pps(tau);N=50000;nrep=3;nbpts=100;
    fic="test7";verifvp=function(y){(((y>0)*(y<2)/2)/(N*b(N))*(y/tau +y^2))/(2
    *sqrt(pi))};
sim6<-Verif(m,N,b,nrep,nbpts,fic,verifvp);
save(sim6,file="sim7")

(m$dloi(y0)/(N*b(N)))*(m$vinf(y0)/(m$tau^2)+((m$rho(y0))^2))*ker$intK2
```

```
##5. Chapter 5 Pseudo maximum likelihood - stratified sampling and linear
    model
#5.1. Calculus of the weight function
rhoh<-function(y,proph,tauh,theta,xi,sigma){
         rhorho<-tauh[length(tauh)];t<-proph[1];tau=sum(proph*tauh)
       for(h in 1:(length(tauh)-1)){rhorho<-rhorho+(tauh[h]-tauh[h+1])*pnorm((
           sqrt(xi^2+sigma^2)*qnorm(t)+xi*(theta-y))/sigma);t<-t+proph[h+1]}
       return(rhorho/tau)}
##5.2. Calculus of the mean sample log likelihood
sample.like<-function(theta,y,m,xi.hat=xi.hat){
    W<-model.dep.strat(proph=m$param$proph,tauh=m$param$tauh,theta=theta,xi=xi
        .hat,sigma=m$param$sigma)$rho(y)
    f<-model.dep.strat(proph=m$param$proph,tauh=m$param$tauh,theta=theta,xi=xi
        .hat,sigma=m$param$sigma)$dloi(y)
    log.like<-sum(log(W)+log(f))
    mlog.like<-  -log.like
return(mlog.like)
}
##5.3. Calculus of differentiate(Delta)
    deriveetheta1<-function(y,proph,tauh,theta,xi,sigma){
        H<-length(tauh);       deno<-tauh[H];nume<-0
        t=proph[1]
        for(h in (1:(H-1))){
          deno<-deno+(tauh[h]-tauh[h+1])*pnorm((sqrt(xi^2+sigma^2)*qnorm(t)+xi
              *(theta-y))/sigma,mean=0,sd=1);
          nume<-nume+(tauh[h]-tauh[h+1])*(xi/sigma)*dnorm((sqrt(xi^2+sigma^2)*
              qnorm(t)+xi*(theta-y))/sigma,mean=0,sd=1)
          t<-t+proph[h+1]}
          return((y-theta)+(nume/deno))}
    deriveetheta<-function(y,proph,tauh,theta,xi,sigma){
          return(sapply(y,deriveetheta1,proph=proph,tauh=tauh,theta=theta,xi=
              xi,sigma=sigma))}
    deriveexi1<-function(y,proph,tauh,theta,xi,sigma){
        H<-length(tauh);       deno<-tauh[H];nume<-0
        t=proph[1]
        for(h in (1:(H-1))){
          deno<-deno+(tauh[h]-tauh[h+1])*pnorm((sqrt(xi^2+sigma^2)*qnorm(t)+xi
              *(theta-y))/sigma,mean=0,sd=1);
          nume<-nume+(tauh[h]-tauh[h+1])*(((xi/sigma^2)*((xi^2/sigma^2+1)^(-1/
              2))*qnorm(t)+(theta-y))/sigma)*dnorm((sqrt(xi^2+sigma^2)*qnorm(t
              )+xi*(theta-y))/sigma,mean=0,sd=1)
          t<-t+proph[h+1]}
          return((nume/deno))}
    deriveexi<-function(y,proph,tauh,theta,xi,sigma){
          return(sapply(y,deriveexi1,proph=proph,tauh=tauh,theta=theta,xi=xi,
              sigma=sigma))}
##5.4. Calculus of asymtotic variance covariance matrix $\Sigma$
    nrep<-10;theta<-1.5;xi<-2;sigma<-1;proph=c(.7,.3);tauh=c(1/70,2/15);N<-
        5000
      cav<-function(proph,tauh,theta,xi,sigma,nrep){

        H=length(tauh);th=vector()
        Bh=vector();th[1]=proph[1];
```

```
        for(h in 2:H){th[h]=proph[h-1]+proph[h]}
        Bh=xi*theta+sqrt(xi^2+sigma^2)*qnorm(th)
        toto<-function(z,Bh){(z>c(-Inf,Bh[-H]))*(z<=Bh)}
        y=rnorm(nrep,mean=theta,sd=1);
        eta=rnorm(nrep,mean=0,sd=sigma);
        z=xi*y+eta
        dd= deriveetheta(y,proph=proph,tauh=tauh,theta=theta,xi=xi,sigma=sigma
            )
        ddxi= deriveexi(y,proph=proph,tauh=tauh,theta=theta,xi=xi,sigma=sigma)
        rhorho=model.dep.strat(proph,tauh,theta,xi,sigma)$rho(y)
        ind<-sapply(z,toto,Bh=Bh);
        ind[ind==0]<-NA
        w<-ind/tauh
        ind<-t(ind)   ;w<-t(w)

        Vh.ypi_1<-apply(y*w,2,var,na.rm=TRUE)
        Vh.y<-apply(y*ind,2,var,na.rm=TRUE)
        Vh.y2pi_1<-apply(y^2*w,2,var,na.rm=TRUE)
        Vh.yzpi_1<-apply(y*z*w,2,var,na.rm=TRUE)
        Vh.delta<-apply(dd*ind,2,var,na.rm=TRUE)
        Covh.y2pi_1yzpi_1<-apply((y^3*z)*w^2,2,mean,na.rm=TRUE)-apply(y^2*w,2,
            mean,na.rm=TRUE)*apply(y*z*w,2,mean,na.rm=TRUE)
        Covh.y2pi_1delta<-apply(y^2*dd*w,2,mean,na.rm=TRUE)-apply(y^2*w,2,mean
            ,na.rm=TRUE)*apply(dd*ind,2,mean,na.rm=TRUE)
        Covh.yzpi_1delta<-apply(y*z*dd*w,2,mean,na.rm=TRUE)-apply(y*z*w,2,mean
            ,na.rm=TRUE)*apply(dd*ind,2,mean,na.rm=TRUE)
        I11=mean(dd^2*rhorho)
        I12=mean(dd*ddxi*rhorho)
        Sigmah<-list();
        for(h in 1:H){
          Sigmah[[h]]<-
            matrix(c(Vh.delta[h],Covh.yzpi_1delta[h],Covh.y2pi_1delta[h],Covh.
                yzpi_1delta[h],Vh.yzpi_1[h],Covh.y2pi_1yzpi_1[h],Covh.y2pi_1
                delta[h],Covh.y2pi_1yzpi_1[h],Vh.y2pi_1[h]),3,3)}
        Sigma<-0 ;
        for(i in (1:H)){
          Sigma<-tauh[h]*Sigmah[[h]]*proph[h]+Sigma}
        MA<-matrix(c(1,0,0,0,1/(theta^2+1),-xi/(theta^2+1)),3,2)
        Sigma<-1/(sum(tauh*proph))*t(MA)%*%Sigma%*%MA
        V<-Sigma[1,1]/(I11^2)+I12*(Sigma[2,2]*I12-2*Sigma[1,2])/I11^2
         V1<-Sigma[1,1]/(I11^2)
          VHT<-(sum(tauh*proph))*sum(tauh*proph*Vh.ypi_1)
           Vniais<-1/(sum(tauh*proph))*sum(tauh*proph*Vh.y)
         return(list(Sigma=Sigma,I11=I11,I12=I12,VHT=VHT,V=V,V1=V1,Vniais=
             Vniais))}
##5.4.1.
affiche<-function(x){puis<-floor(log(x)/log(10));
af<-paste("",signif(10^(-puis)*x,4),"\\ 10^{",puis,"}");
if(puis>=0&puis<4){af<-paste("",signif(x,4),sep='')}
return(af)}
##5.5. Simulation procedure.
simule<-function(N,m,nbreps){
  #initialization
```

```
    theta.hat<-rep(0,nbreps);theta.tilde<-rep(0,nbreps);theta.ht<-rep(0,nbreps
        );xi.hat<-rep(0,nbreps);xi.ht<-rep(0,nbreps);theta.breve<-rep(0,nbreps
        );theta.bar<-rep(0,nbreps)

  for(i in 1:nbreps){
    Yg<-m$rloiy(N); #Y generation
    Zg<-m$rloiz(Yg); #Z generation
    pik<-m$Scheme$Pik(Zg); #inclusion probabilities
    Sg<-m$Scheme$S(Zg); #sample selection
    s.zy<-sum((Zg*Yg/pik)[Sg]) #HT estimator of $\sum_{k=1}^N Y_k Z_k$
    s.y2<-sum((Yg^2/pik)[Sg])  #HT estimator of $\sum_{k=1}^N Y_k^2$
    s.z<-sum((Zg/pik)[Sg])     #HT estimator of $\sum_{k=1}^N Z_k$
    s.y<-sum((Yg/pik)[Sg])     #HT estimator of $\sum_{k=1}^N Y_k$
    s.1<-sum((1/pik)[Sg])      #HT estimator of $N$
    xi.ht[i]<-(s.zy)/(s.y2)    #estimator of $\xi$
    ys<-Yg[Sg]                 #vector of sample responses
    #calculus of three different estimators of theta
    theta.hat[i]<-optimize(f=sample.like,interval=c(0,5),y=ys,m=m,xi=xi.ht[i])
        $minimum
    theta.bar[i]<-mean(ys)
    theta.ht[i]<-s.y/s.1}
    return(list(
         cebon=sqrt(mean((theta.ht-theta)^2)/mean((theta.hat-theta)^2)),
         xi.ht=xi.ht,
         theta.hat=theta.hat,
         theta.bar=theta.bar,
         theta.ht=theta.ht,
         m.hat=mean(theta.hat),
         m.ht=mean(theta.ht),
         m.bar=mean(theta.bar),
         mse.hat=mean((theta.hat-theta)^2),
         mse.ht=mean((theta.ht-theta)^2),
         mse.bar=mean((theta.bar-theta)^2)))}

##5.6 procedure that launches simulations and produces an output:
##    a tex code for a table containing the results of simulation

generetableau<-function(nbreps,tauh,proph,N,Theta,Xi,Sigma,fic){
  write("\\begin{tabular}{|ccc|c|r|r|r|r|}",file=fic,append=F)
  write("\\hline",file=fic,append=T)
  write("$\\theta$& $\\xi$& $\\sigma$
      &
      &Mean.
      &M.S.E.
      &$\\sqrt{\\frac{\\rm{MSE}}{\\rm{MSE}(\\hat{\\theta})}}$
      &$\\frac{1}{n_{\\gamma}}\\lim_{\\gamma\\to\\infty}n_{\\gamma}\\V{.}$
          \\\\\\hline",file=fic,append=T)
  for(k in 1:length(Theta)){
  tau<-sum(proph*tauh)
  theta<-Theta[k];xi<-Xi[k];sigma<-Sigma[k];
    m<-model.dep.strat(proph=proph,tauh=tauh,theta=theta,xi=xi,sigma=sigma)
    sim=simule(N=N,m,nbreps=nbreps)
    save(sim,file=paste("sim_",k,sep=''))
```

```
      cave<-cav(proph=c(.7,.3),tauh=c(1/70,2/15),theta=theta,xi-xi,sigma=sigma,
          nrep=60000)
       save(cave,file=paste("cave_",k,sep=''))
   write(paste("$",Theta[k],"$& $",Xi[k],"$& $",Sigma[k]  ,"$
         &$\\hat{\\theta}$
         &$",affiche(sim$m.hat),"$
         &$",affiche(sim$mse.hat),"$&$1$
         &$",affiche(cave$V/(N*tau)),"$
         \\\\"),file=fic,append=T)
   write(paste("&&
         &$\\tilde{\\theta}$
         &$",affiche(sim$m.ht),"$
         &$",affiche(sim$mse.ht),"$
         &$",affiche(sqrt(sim$mse.ht/sim$mse.hat)),"$
         &$",affiche(cave$VHT/(N*tau)),"$ \\\\"),file=fic,append=T)
   write(paste("& &
         &$\\bar{\\theta}$
         &$",affiche(sim$m.bar),"$
         &$",affiche(sim$mse.bar),"$
         &$",affiche(sqrt(sim$mse.bar/sim$mse.hat)),"$
         &$",affiche(cave$Vniais/(N*tau)),"$ \\\\\\\\hline"),file=fic,append=T)}
   write("\\end{tabular}",file=fic,append=T)}
generetableau(nbreps=30,tauh=c(1/70,2/15),proph=c(.7,.3),N=1000,Theta=c
    (1.5,1.5,1.5),Xi=c(2,2,2),Sigma=c(0.1,1,10),fic="graph7.tex")

##5.7. Graphs for chapter 5

maxy=max(Gg$Yg);maxz=max(Gg$Zg);  fic3="figessai"
prop1<-c(-0.1,-0.12,1.05,1.2);prop2<-c(-0.2,-0.3,1.05,1.2);
  maxz3=1
N=500;theta<-1.5;xi<-2;y<-rnorm(N,mean=theta);proph<-c(.7,.3);tauh<-c(.01,.08)
    ;
#graphics for chapter5
m<-model.dep.strat(proph,tauh,xi,theta,sigma=5)
Gg<-genere(m=m,N=N,y);maxy=.95*max(Gg$Yg);maxz=max(Gg$Zg)
plotgen("figessai3",Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,
    listcouleur=listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,maxz3,
    quoigen=c(1))
m<-model.dep.strat(proph,tauh,xi,theta,sigma=.1)
Gg<-genere(m=m,N=N,y);maxz=.85*max(Gg$Zg)
plotgen("figessai",Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,
    listcouleur=listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,1,
    quoigen=c(1))
m<-model.dep.strat(proph,tauh,xi,theta,sigma=1)
Gg<-genere(m=m,N=N,y);maxz=.85*max(Gg$Zg)
plotgen("figessai2",Gg,b,ker=ker,m=m,prop1,prop2,taillegraph,demarc=NULL,
    listcouleur=listcouleur,listtaille=listtaille,repy,repz,maxy,maxz,maxz3,
    quoigen=c(1))
```

General Bibliography

Arratia, R., Goldstein, L., and Langholz, B. (2005). Local central limit theorems, the high-order correlations of rejective sampling and logistic likelihood asymptotics. *The Annals of Statistics*, 33(2):871–914.

Berger, Y. G., Muñoz, J. F., and Rancourt, E. (2009). Variance estimation of survey estimates calibrated on estimated control totals– An application to the extended regression estimator and the regression composite estimator. *Computational Statistics & Data Analysis*, 53(7):2596 – 2604.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *Journal of Statistical Planning and Inference*, 51(3):279–292.

Bonnéry, D., Breidt, F. J., and Coquet, F. (2011). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli*. To appear.

Bonnéry, D., Chauvet, G., and Deville, J.-C. (2009). Optimum de type neyman pour l'échantillonnage équilibré sur des marges. In *Actes des Journées de Méthodologie Statistique*.

Breckling, J., Chambers, R. L., Dorfman, A. H., Tam, S., and Welsh, A. (1994). Maximum likelihood inference from sample survey data. *International Statistical Review/Revue Internationale de Statistique*, 62(3):349–363.

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *The Annals of Statistics*, 28(4):1026–1053.

Breidt, F. J. and Opsomer, J. D. (2008). Endogenous post-stratification in surveys: classifying with a sample-fitted model. *The Annals of Statistics*, 36(1):403–427.

Cassel, C.-M., Särndal, C.-E., and Wretman, J. H. (1977). *Foundations of Inference in Survey Sampling*. Wiley-Interscience [John Wiley & Sons], New York. Wiley Series in Probability and Mathematical Statistics.

Chambers, R. L., Dorfman, A. H., and Wang, S. (1998). Limited information likelihood analysis of survey data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(2):397–411.

Chambers, R. L. and Skinner, C. J. (2003). *Analysis of Survey Data*. Wiley Series in Survey Methodology. John Wiley & Sons Inc, Chichester.

Chauvet, G., Bonnéry, D., and Deville, J.-C. (2011). Optimal inclusion probabilities for balanced sampling. *Journal of Statistical Planning and Inference*, 141(2):984–994.

Chen, X., Dempster, A., and Liu, J. S. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3):457.

Cox, D. (1969). *Some Sampling Problems in Technology*, pages 506–527. Wiley Interscience.

Deville, J.-C. (2000). Note sur l'algorithme de chen, dempster et liu. *Rapport technique, CREST-ENSAI, Rennes*.

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4):893–912.

Deville, J.-C. and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128(2):569–591.

Eideh, A. A. H. and Nathan, G. (2006). Fitting time series models for longitudinal survey data under informative sampling. *Journal of Statistical Planning and Inference*, 136(9):3052–3069.

Eideh, A. A. H. and Nathan, G. (2007). Corrigendum to "fitting time series models for longitudinal survey data under informative sampling". *Journal of Statistical Planning and Inference*, 137(2):628.

Eideh, A. A. H. and Nathan, G. (2009). Two-stage informative cluster sampling-estimation and prediction with applications for small-area models. *Journal of Statistical Planning and Inference*, 139(9):3088–3101.

Fuller, W. A. (2009). *Sampling Statistics*. John Wiley & Sons Inc.

Gong, G. and Samaniego, F. J. (1981). Pseudomaximum likelihood estimation: theory and applications. *The Annals of Statistics*, 9(4):861–869.

Gourieroux, C. (1981). *Théorie des Sondages*. Economica.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35(4):1491–1523.

Hájek, J. (1981). *Sampling from a Finite Population*, volume 37 of *Statistics: Textbooks and Monographs*. Marcel Dekker Inc., New York. Edited by Václav Dupač, With a foreword by P. K. Sen.

Hausman, J. and Wise, D. (1981). *Stratification on endogenous variables and estimation: The Gary income maintenance experiment*, pages 36–391. Cambridge, MA: MIT Press.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77(377):89–96.

Kish, L. and Frankel, M. (1974). Inference from complex surveys. *Journal of the Royal Statistical Society, Series B*, 36(1):1–37.

Langholz, B. and Goldstein, L. (2001). Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics*, 2(1):63–84.

Leigh, G. M. (1988). A comparison of estimates of natural mortality from fish tagging experiments. *Biometrika*, 75(2):347–353.

Mantel, N. (1973). Synthetic retrospective studies and related topics. *Biometrics*, 29:479–486.

Matei, A. and Tillé, Y. (2005). Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size. *Journal of Official Statistics*, 21(4):543.

Nowell, C. and Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research*, 28(4):475–479.

Patil, G. P. and Rao, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics*, 34(2):179–189.

Pfeffermann, D. and Krieger, A. M. (1992). Maximum likelihood estimation for complex sample surveys. *Survey Methodology*, 18(2):225–239.

Pfeffermann, D., Krieger, A. M., and Rinott, Y. (1998). Parametric distributions of complex survey data under informative probability sampling. *Statistica Sinica*, 8(4):1087–1114.

Pfeffermann, D., Moura, F. A. D. S., and do Nascimento Silva, P. L. (2006). Multi-level modelling under informative sampling. *Biometrika*, 93(4):943–959.

Pfeffermann, D. and Sverchkov, M. Y. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhyā: The Indian Journal of Statistics, Series B*, 61(1):166–186.

Pfeffermann, D. and Sverchkov, M. Y. (2003). Fitting generalized linear models under informative sampling. In *Analysis of survey data*, Wiley Series in Survey Methodology, pages 175–195. Wiley, Chichester.

Pfeffermann, D. and Sverchkov, M. Y. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association*, 102(480):1427–1439.

Pfeffermann, D. and Sverchkov, M. Y. (2009). Inference under Informative Sampling. In Pfefferman, D. and Rao, C., editors, *Sample Surveys: Inference and Analysis*, volume 29B of *Handbook of Statistics*, pages 455–487. Elsevier.

Robinson, P. and Särndal, C.-E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhyā: The Indian Journal of Statistics, Series B*, 45(2):240–248.

Särndal, C.-E., Swensson, B., and Wretman, J. H. (1992). *Model Assisted Survey Sampling*. Springer Series in Statistics. Springer-Verlag, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Shaw, D. (1988). On-site samples' regression: problems of nonnegative integers, truncation, and endogenous stratification. *Journal of Econometrics*, 37(2):211–223.

Skinner, C. J. (1994). Sample models and weights. In statistical association, A., editor, *Proceedings of the Section on Survey Research Methods*, pages 133–142, Washington, DC.

Snyman, J. (2005). *Practical Mathematical Optimization: an Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, volume 97. Springer Verlag.

Sullivan, P., Breidt, F. J., Ditton, R., Knuth, B., Leaman, B., O'Connell, V., Parsons, G., Pollock, K., Smith, S., and S.Stokes (2006). *Review of Recreational Fisheries Survey Methods*. National Academies Press, Washington, DC.

Tillé, Y. and Favre, A. (2005). Optimal allocation in balanced sampling. *Statistics & Probability Letters*, 74(1):31–37.

Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer Verlag.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Index of terms

cardinal number, 83
cdf, cumulative distribution function
  empirical population cdf, 40
cdf,cumulative distribution function
  limit sample cdf, 28
  population cdf, 28
characteristic, 1

design measure, 13, 14, 23, 66
  design measure function, 14
design variable, 23
design variables, 1, 14

element, 1
exchangeable, 84

Hájek, 18
Horvitz Thompson, 18

inference
  design-based inference, 1, 17
  model-based inference, 1, 18
informative, non informative selection, 1
informative, non-informative selection
  under the exchangeable population model, 20
  under the iid superpopulation model, 20
  under the fixed population model, 21

kernel, 45

label
  labelled, unlabelled, 15

measure
  counting measure, 84
  Dirac measure, 83
  uniform probability measure, 84

observation, 16, 23
  statistical model for observations
    design-based model, 17

population, 1, 13, 22
  population model, 2
    exchangeable model for fixed population, 19
    exchangeable population model, 18
    fixed population model, 17
    superpopulation model, 1
  population size, 1, 13
population model
  iid superpopulation model, 18
power set, 83

sample, 1, 13, 15, 66
  limit sample distribution, 25
  limit sample probability density function, 25
  sample distribution, 23
  sample empirical cumulative density funtion, 2
  sample empirical distribution, 24
  sample probability density function, 2, 23
  sample size, 15, 23
sampling
  balanced sampling, 3
  Bernoulli sampling, 32
  Poisson sampling, 14, 33
  simple random sampling, 13
  with replacement, 14
  without replacement, 14
sampling design, 1
  cluster sampling, 54
study variables, 1, 14

Index of notations

!, 83
#, 83
$\binom{N}{n}$, 83
$\lfloor . \rfloor$, 84
Pois_z, 33
$[.,.]$, $].,.]$, $[.,.[$, $].,.[$, 83
$\sim$, 84

$\mathbb{1}_A$, 83

$\mathscr{A}$, σ-field on Ω, 14
A^B, set of functions from A to B, 83

$\mathrm{Bern}_{N,p}$, Bernoulli sampling, 32

C, c, 20
c_γ, 27
$c_{\gamma,\theta,\xi}$, 58
$\mathscr{C}^l$, 83

D, design measure function, 14
Δ, 59
δ, 84
D_γ, design measure function, 23
d_γ, 27
$d_{\gamma,\theta,\xi}$, 58

F, population cdf, 28
$\mathscr{F}(A,B)$, set of functions from a set A to a set B, 83
F_γ, empirical sample cdf, 27
f_γ, density of $\mathrm{P}_{\gamma k}^Y$ w.r.t. μ_Y, 23
$\hat{f}_\gamma$, 53
F_∞, limit sample cdf, 28
$f.\mu$, 84

G, the observations, 16
g, observation function, 16
γ, population index, 22

h_γ, a bandwidth, 45

$\mathcal{I}$, a random sample, 14, 15
i, a sample, 13, 66
$\mathscr{I}_{11}$, 60
$\mathscr{I}_{12}$, 60
$\mathrm{Im}(f)$, image set of f, 83

Inj, 83
$\mathcal{I}^*$, 16

K, a kernel, 45
k, an element, 13, 66

λ, the Lebesgue measure, 27

M, 24
m'_γ, 27
m_γ, 27
$m'_{\gamma,\theta,\xi}$, 58
m_∞, 24
$m_{\infty,\theta,\xi}$, 58
μ_Y^*, measure on $(\mathscr{Y}, \mathscr{T}_Y)$, 24
μ_Y, Radon measure on $(\mathscr{Y}, \mathscr{T}_Y)$, 23

N, population size, 13
$\mathscr{N}(.,.)$, the normal distribution, 84
$\mathbb{N}$, 83
N_γ, population size, 22

Ω, 14
$(\Omega_p, \mathscr{A}_p, \mathrm{P}_p)$, $(\Omega_s, \mathscr{A}_s, \mathrm{P}_s)$, $(\Omega_l, \mathscr{A}_l, \mathrm{P}_l)$, probability spaces, 15
o_P ,O_P, 84

P, probability on $(\Omega, \mathscr{A})$, 14
P^Y, 84
$\mathscr{P}$, 83
p, a design measure, 13, 66
$\mathcal{P}(\beta, L)$, 45
Φ, the cumulative density function of the normal distribution, 84
ϕ, the density function of the normal distribution, 84
Π, random design measure on U, 14
Π_γ, 23
$\mathbb{\Pi}$, set of design measures, 13

Q, 19
$\mathbb{Q}$, 83

R, 15
$r.a$, 83
$\mathbb{R}$, 83

R_γ, random labelling of observations in the sample from U_γ, 23
ρ, 2
ρ_γ, weight function, 23
$\rho_{\gamma,\theta,\xi}$, 58
ρ_∞, 25
$r\tilde{h}o_\infty$, 53

$\Sigma(\beta, L, T)$, 45
$\mathfrak{S}_N$,set of permutations of $\{1, \ldots, N\}$, 83
$\mathrm{SRS}_{N,n}$, simple random sampling of n elements from N elements, 14
SWR_{z,n^*} Sampling with replacement of size n^* with probability proportional to z, 38

θ, 3
$\hat{\theta}_\gamma(\xi)$, 59
$\mathscr{T}_Y$, σ-field on $\mathscr{Y}$, 14
$\mathscr{T}_Z$, σ-field on $\mathscr{Z}$, 14

U, population, 13
U_γ, γth population , 22

v_γ, 27
$v_{\gamma,\theta,\xi}$, 58
v_∞, 50

ξ, 3
X_p, X_s, X_l projections on Ω_p, Ω_s, Ω_l, 16

$\mathcal{Y}$, vector of study variables, 14
y, vector of study variable in the design-based case, 3
$\mathscr{Y}$, 14
$\mathcal{Y}_\gamma$, vector of study variables for population U_γ, 22
$Y_{\gamma k}$, study variables for element k of U_γ, 22
Y_k, 14, 27
$\mathcal{Y}^*$, 16

$\mathcal{Z}$, 14
$\mathscr{Z}$, 14
$\mathbb{Z}$, 83
$\zeta_\alpha(z)$, 34, 54
$\mathcal{Z}_\gamma$, vector of design variables for U_γ, 22
$Z_{\gamma k}$, value of design variable for element k of U_γ, 22
Z_k, 14
$\mathcal{Z}^*$, 16

List of Figures

2.1 Commutative diagram for $\mathcal{Y}, \mathcal{Z}, \Pi, \mathcal{I}, R$ . . . 18
2.2 Commutative diagram for $\mathcal{Y}_\gamma, \mathcal{Z}_\gamma, \Pi_\gamma, \mathcal{I}_\gamma, R_\gamma$ . . . 25

3.1 Cluster sampling example showing population cdf, limit sample cdf, empirical population cdf, and empirical sample cdf. . . . 38
3.2 With-replacement sampling with probability proportional to size . . . 42
3.3 With-replacement sampling with probability proportional to size and Cauchy distribution . . . 43

4.1 Independent stratified sampling . . . 56
4.2 Kernel density estimation in the case of cluster sampling . . . 57
4.3 Sampling with replacement and probability proportional to size . . . 58

5.1 Different degrees of informative selection . . . 65

List of Tables

5.1 Mean square error of different estimators . . . 66

6.1 Exact quantities needed for the computation of optimal inclusion probabilities with Algoritm 6.1, for the vectors $\sigma^{(1)}$ and $\sigma^{(2)}$ . . . 78
6.2 Three sets of inclusion probabilities obtained with the fixed-point algorithm for three variables of interest, for the vectors $\sigma^{(1)}$ and $\sigma^{(2)}$ . . . 78
6.3 Total of the variables of interest and variance approximation for three sets of inclusion probabilities . . . 79
6.4 Relative Efficiency of the optimal vector of inclusion probabilities . . . 79
6.5 Comparison of Algoritm 6.1 Tillé and Favre method . . . 81
6.6 Relative Efficiency when prior information is available . . . 81

List of Algorithms

6.1 Fixed-point algorithm to compute optimal inclusion probabilities . . . 74
6.2 Fixed-point algorithm to compute approximately optimal inclusion probabilities . . . 76
6.3 Gradient descent to determine local optimal inclusion probabilities . . . 83

VU :
Le Directeur de Thèse
François COQUET

VU :
Le Responsable de l'École Doctorale

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication :

Le Président de Jury,
(Nom et Prénom)