The Informational Order in Ranked Set Sampling Experiments

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
Denise Steliana Rey
aus
Bukarest

Göttingen 2004

D7
Referent: Prof. Dr. M. Denker
Korreferent: Prof. Dr. A. Munk
Tag der mündlichen Prüfung: 28.04.04

Contents

1 Introduction

The main content of this dissertation is the comparison of statistical experiments which are based on the Ranked Set Sampling (RSS) technique. Using the randomization criterion in the theory of experiments we decide on the existence or non-existence of an informational order between RSS experiments.

Sampling and Resampling Techniques

In the era of cheap and fast computation, more and more attention is given to sampling (resampling respectively) problems. The purpose of sampling (resampling) theory is to make sampling (resampling) of observations more efficient with respect to *costs*, *speed* of sampling (resampling) and *information* gained from a sample, which is used in further statistical inference. From an applied point of view, sampling techniques are defined before the process of observing and measuring and then directly involved into it, while resampling techniques is a generic name for all methods which evaluate an estimator of a parameter with the help of reweighted versions of the empirical probability distribution, i.e. resampling techniques rely on an already existing sample of observations. Cochran's [10] book *Sampling Techniques* is one of the first books giving an overview of the most important sampling procedures used in sampling surveys, like *Simple Random Sampling* (further denoted by SRS), *Stratified Random Sampling*, *Systematic Sampling*, *Double Sampling*, etc. Quenuille [22] and later Tukey [28] invented a nonparametric estimate of the bias respectively the variance, subsequently named the *Jackknife*. The *Bootstrap* method, first mentioned by Efron [13] is a more general method of estimation. The latter two method are as the *Random Subsampling*, the *balanced Repeated Replications*, resampling techniques.

The Ranked Set Sampling

The technique *Ranked Set Sampling* (RSS), a sampling technique, was first introduced by McIntyre [20] as an efficient alternative to simple random sampling for estimating the expected pasture yields. In environmental and ecological sampling, or more generally in spacial sampling, one may encounter situations where exact measurements of the variable of interest are expensive (in terms of time, money, or other), but where ranking on the basis of visual inspection or on the basis of another highly correlated random variable can be done easily. In McIntyre's case, measuring the plots of pasture yields requires moving and weighting crop yields, which is time consuming. However, a small number of plots can be even though sufficiently well ranked by eye without measurement. McIntyre's goal was to develop a sampling technique to reduce the number of necessary measurements to be made, maintaining the unbiasedness of the SRS mean and reducing the variance of the mean estimator by incorporating the outside information provided by visual inspection. Therefore, since the ranking of the plots could be done very cheap, he developed a technique to implement this advantage.

In the original form, the practical RSS technique can be described as follows: A set of p, $p \geq 2$, independent and identically distributed (i.i.d.) random variables X_i, $1 \leq i \leq p$, is drawn from a population with unknown distribution, denote it by P^X. In practice, the items of the set are not yet measured, but ranked visually. The item which is believed to be the smallest is measured, denote it by $X_{[1]}$. Then another set p of i.i.d. random variables is drawn, ranked visually and the item which is believed to be the second smallest is measured, denote it by $X_{[2]}$. The procedure is repeated until the item ranked the largest in the p-th set of p i.i.d. random variables is measured, say $X_{[p]}$. This completes a cycle of the sampling and is called in literature the RSS technique with one cycle. McIntyre estimates the expectation of the underlying probability distribution P^X with the RSS mean estimator $p^{-1} \sum_{i=1}^{p} X_{[i]}$ and concludes that it is unbiased and has a smaller variance than the sample mean estimator obtained via SRS, $p^{-1} \sum_{i=1}^{p} X_i$. In practice, the set size p is kept small to ease the visual ranking and the cycle mentioned above can be repeated several times. The balanced RSS technique with n cycles then gives rise to the random variables $X_{i[1]}, \ldots, X_{i[p]}$, $1 \leq i \leq n$.

McIntyre's RSS technique did not find further interest for over a decade. Then Halls [14] conducted a field trial by using RSS technique for estimating forage yields in a pine forest. They reported the gain of efficiency by using RSS instead of SRS and mentioned also the practical problems which can arise when using RSS. Halls gave also the name *Ranked Set Sampling* which is today in use. The first attempts of a mathematical modeling of the RSS technique were made by Takahasi [27] and Dell [12]. They showed that the relative efficiency of the RSS mean estimator (in the perfect case, to be defined later) with respect to the SRS mean estimator is bounded below by 1 and above by $\frac{p+1}{2}$ for all continuous distributions with a finite variance. Since then, the RSS technique has found more and more interest in diverse statistical problems and a considerable number of papers have been published. These papers can be classified into theoretical papers, design improvement papers and practical applications papers.

For an overview of the papers dealing with theoretical statistical problems, we introduce some notations: Assume $X_1, \ldots, X_{np}$ is a SRS sample of size np, $p \geq 2$, $n \in \mathbb{N}$, of i.i.d. random variables with distribution P^X. Further, let $X_{i[1]}, \ldots, X_{i[p]}$, $1 \leq i \leq n$, be independent random variables, $X_{i[j]} \sim P^{X_{[j]}}$ for all $1 \leq i \leq n$, $1 \leq j \leq p$, the RSS with n cycles. We do not give here a detailed description of the distribution of the RSS random variables. We only mention the difference between the *perfect* RSS and *imperfect* RSS, which has found great attention in the literature. Loosely speaking, the perfect RSS is the RSS technique in case the visual ranking is done perfectly, i.e. the j-th measurement in the j-th sample is with probability 1 the j-th largest measurement in a sample of size p. Theoretically this means $P^{X_{[j]}} := P^{X_{(j:p)}}$ is the distribution of the j-th order statistic in a sample of size p and in this case we denote $X_{i[j]} := X_{i(j)}$, $1 \leq i \leq n$ and $1 \leq j \leq p$. All the other cases are called in literature *imperfect* RSS, *error in model* RSS or *judgment* RSS. In this dissertation, we give a well defined mathematical expression for the above mentioned cases of RSS.

In what follows we list some of the relevant statistical decision problems which were treated by the RSS technique and mention the important results.

- Let the functional to be estimated be expectation $E_{P^X}X$. This estimation problem was first treated by McIntyre [20]. The SRS estimator compared with the RSS estimator is the usual sample mean, $\bar{X}_{SRS} = \frac{1}{np}\sum_{i=1}^{np} X_i$. The RSS estimator is defined by $\bar{X}_{RSS} := \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p} X_{i[j]}$. The RSS estimator is unbiased: $E[\bar{X}_{RSS}] = E_{P^X}X$. Moreover it has an asymptotic normal distribution, $\sqrt{np}(\bar{X}_{RSS} - E_{P^X}X) \overset{n\to\infty}{\to} N(0, \frac{1}{p}\sum_{j=1}^{p}\text{var}[(X_{1[j]})])$. The gain of information by using RSS as compared to SRS, in terms of relative efficiency is obtained as

$$RE := \frac{\text{var}[\bar{X}_{SRS}]}{\text{var}[\bar{X}_{RSS}]} = 1 + \frac{\sum_{j=1}^{p}(E[X_{1[j]}] - E_{P^X}X)^2}{\sum_{j=1}^{p}\text{var}[X_{1[j]}]} \geq 1.$$

- Let the functional to be estimated be the variance $\text{var}_{P^X}[X]$. This estimation problem was first treated in the case of perfect RSS by Stokes [24]. The SRS estimator considered is $s^2_{SRS} = (np-1)^{-1}\sum_{i=1}^{np}(X_i - \bar{X}_{SRS})^2$. The RSS Estimator is $s^2_{RSS} = (np-1)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p}(X_{i(j)} - \bar{X}_{RSS})^2$. The RSS estimator is biased, $E[s^2_{RSS}] = \text{var}_{P^X}[X] + \frac{\sum_{j=1}^{p}(EX_{1(j)} - EX)^2}{p(np-1)}$ but asymptotically unbiased as $\lim_{n\to\infty} E[s^2_{RSS}] = \text{var}_{P^X}[X]$. The gain of information in terms of asymptotic relative efficiency can be summarized as follows: $\exists N \in \mathbb{N}$ such that for $np > N$, $\frac{\text{var}[s^2_{SRS}]}{\text{var}[s^2_{RSS}]} \geq 1$. For example, if $P^X = N(0,1)$ then we need $p \geq 5$ to achieve the latter inequality. The asymptotic relative efficiency is then $ARE = \lim_{n\to\infty}\frac{\text{var}[s^2_{SRS}]}{\text{var}[s^2_{RSS}]} \geq 1$.

- A more recent application is the kernel estimation problem, as treated for example also in the case of perfect RSS by Barabesi [2]. Consider $P^X \ll \lambda$ where we have denoted by λ the Lebesgue measure on the real line. Then the functional to be estimated is $f = \frac{dP^X}{d\lambda}$. The usually SRS kernel estimator is $\hat{f}_{SRS}(x) := (np)^{-1}\sum_{i=1}^{np} K_h(x - X_i)$, $K_h(u) = h^{-1}K(u/h)$ where K is a kernel function. The bandwidth h is chosen such that consistency of the estimator is achieved. The RSS estimator as given by Barabesi is $\hat{f}_{RSS}(x) := (np)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{p} K_h(x - X_{i(j)})$. The expectation of the RSS estimator equals the expectation of the SRS estimator: $E[\hat{f}_{SRS}(x)] = E[\hat{f}_{RSS}(x)] = EK_h(x - X)$. The two estimators are asymptotically equivalent, therefore in the case of kernel estimation, the RSS technique is relevant for finite sample sizes. The gain of information by using the RSS technique is described by the relationship

$$\text{var}[\hat{f}_{SRS}(x)] = \text{var}[\hat{f}_{RSS}(x)] + \frac{1}{np^2}\sum_{j=1}^{p}(EK_h(x - X_{1(j)}) - EK_h(x - X))^2,$$

$$MISE(\hat{f}_{SRS}) = MISE(\hat{f}_{RSS}) + \frac{1}{np^2}\sum_{j=1}^{p}\int EK_h(x - X_{1(j)}) - EK_h(x - X))^2 dx,$$

where MISE is the mean integrated squared error.

- The sign test was treated by several authors, under many derivatives of the original RSS technique. We mention here Hettmansperger [15]. Consider the family of probability distributions on the real line $\{P_\theta^X = P^X(x-\theta) : \theta \in \Theta\}$, where θ is the median. Assume $P^X(0) = 1/2$. Consider the test: $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. The usually SRS test statistic is $T_{SRS} := \sum_{i=1}^{np} 1_{\{X_i>0\}}$. The new defined RSS test statistic is denoted by $T_{RSS} := \sum_{i=1}^{n} \sum_{j=1}^{p} 1_{\{X_{i(j)}>0\}}$. Then, under regularity conditions, the Pitman efficiency of the RSS test versus the SRS test is given by

$$e(T_{RSS}, T_{SRS}) = \frac{1}{1 - \frac{4}{p} \sum_{j=1}^{p} (P^{X_{(j)}}(0) - \frac{1}{2})^2} \geq 1.$$

Other relevant problems treated via the RSS technique are for example the comparison of the Fisher information matrix in specific parametric RSS with the Fisher information in specific parametric SRS (Bai [1]), RSS M-estimation for symmetric location families (Zhao [30]), RSS estimation of quantiles (Chen [8]), RSS regression estimation (Chen [9]).

The Theoretical Background

The mathematical tool used for the definition of the RSS technique will be that of the statistical experiment. The notion of a *statistical experiment* was introduced by Blackwell [5] and is now wide-spread. Due to Blackwell and LeCam and in the spirit of Kolmogorov's axiomatic system of the probability theory a statistical experiment is considered a triple for the parameter space $\Theta \neq \emptyset$, $(E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ where E is the sampling space, $\mathcal{B}$ a σ-algebra for E and $\{P_\theta : \theta \in \Theta\}$ is a family of probability measures on $\mathcal{B}$. We therefore define statistical experiments which correspond to the RSS technique and derivatives of it and to SRS. In the defined family of statistical experiments we investigate the relationship between the RSS experiment and the SRS experiment given a decision problem[1] or generally, for every decision problem.

The main difference between the usual decision theory and the theory of statistical experiments is the following: The aim of decision theory is to investigate a *fixed* statistical experiment and to find an optimal decision rule[2] for a given decision problem. Questions of comparison of different statistical experiments for arbitrary decision problems is subject of the general theory of statistical experiments. The main ideas of this theory are formulated by LeCam [19] even though the investigation of the comparison of statistical experiments was initiated by the papers Bohnenblust [7] followed by the papers of Blackwell [5], [6]. A natural question in the theory of experiments is how much "statistical information" contains a considered experiment, or in other words how much information is carried by the observed data. In spite of the various existing definitions of "information" (for example, Fisher information, Shannon information or Kullback information), the definition introduced first in Blackwell [5] and generalized by LeCam [19] is suitable and always defined

[1] A decision problem is a triple (Θ, D, W) consisting of a parameter space Θ, a topological space D of the possible decisions and a loss function W.

[2] Example: For the family of all absolutely continuous w.r.t. Lebesgue measure distribution functions on $\mathbb{R}$, the sample mean as estimator for the expectation has the smallest convex risk in the family of all unbiased estimators.

for a wide class of statistical comparisons. We give here the LeCam version (see the Prerequisites): The statistical experiment $\mathcal{E}$ is more informative than $\mathcal{E}^*$ (denoted by $\mathcal{E}^* \subseteq \mathcal{E}$) for the parameter space Θ, if for every decision (D, W) with a continuous, bounded loss function W and any generalized decision function β^* in the experiment $\mathcal{E}^*$ there exists a generalized decision function β in the experiment $\mathcal{E}$ such that

$$\beta(\mathcal{E}, W) \leq \beta^*(\mathcal{E}^*, W), \quad \theta \in \Theta.$$

For the case of a decision problem on a locally compact space, this definition reduces to the comparison of the risks in the corresponding experiments. There are several methods to compare experiments which lead to assertions equivalent to the definition of information. The criteria to be chosen for the comparison depend mostly on assertions made on the measurable spaces of the experiments as well as on the families of the measures of the experiments. The *randomization* criterion first stipulated by LeCam [19] gives the necessary and sufficient conditions for the existence of an informational order between two arbitrary experiments. The criterion can be summarized as follows: An experiment $\mathcal{E}$ is more informative than another experiment $\mathcal{E}^*$ if and only if $\mathcal{E}^*$ is a randomization of $\mathcal{E}$. If the experiments $\mathcal{E}$ and $\mathcal{E}^*$ are dominated, then $\mathcal{E}^*$ is a randomization of $\mathcal{E}$ if there exists a stochastic operator $M : L_1(\mathcal{E}) \to L_1(\mathcal{E}^*)$ such that $\mathcal{E}^* = M\mathcal{E}$. If the experiment $\mathcal{E}$ is dominated and the measurable space of $\mathcal{E}^*$ is a locally compact separable space, then $\mathcal{E}^*$ is a randomization of $\mathcal{E}$ if there exists a Markov kernel K from $\mathcal{E}$ to $\mathcal{E}^*$ such that $\mathcal{E}^* = K\mathcal{E}$ (for arbitrary experiments this property is called *exhaustivity*). In this dissertation we will need the definition of randomization of two experiments via the existence of Markov kernels.

Motivation of the Comparisons via the Randomization Criterion

Consider the SRS experiment $(\{\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^n P^X : P^X \ll \lambda, \int x^2 P^X(dx) < \infty\})$. Assume we want to estimate the expectation with respect to P^X, i.e. $E_{P^X}[X]$. In this family of probability distributions, the sample mean as a nonrandomized decision function, $\frac{1}{n}\sum_{i=1}^n x_i$ minimizes the convex risk among all mean unbiased estimators. But as we have briefly shown in the examples above, the RSS estimator, constructed also via the sample mean, as the nonrandomized decision function to be taken in the RSS experiment, is unbiased and has a smaller variance than the SRS estimator. Therefore, we do not have to consider the search of an optimal decision in a fixed experiment, but the comparison of the SRS and RSS as statistical experiments. Furthermore, one could search for a sufficient statistic which should induce the RSS experiment and generate via conditional expectation also the sample mean as the nonrandomized decision to be made for the estimation of $E_{P^X}[X]$, but as the results show, there does not exist such a sufficient statistic. Thus, the next step would be not to search for a sufficient statistic, but for a more general exhaustivity relation between SRS and RSS on proper probability spaces.

Results

We give a well-defined expression of the RSS random variables, the perfect and non-perfect cases being differentiated only via distributional assumptions on the model. Assume $\Theta \neq \emptyset$ to be a parameter set, the characteristics of which will be made precise in the thesis. We consider a family of statistical experiments derived by the RSS technique. We only mention here the relevant experiments, the SRS of size n,

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta\}),$$

the RSS of size n without repetition

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} : \theta \in \Theta\}),$$

and the RSS with n repetitions

$$(\mathbb{R}^{np}, \mathcal{B}(\mathbb{R}^{np}), \{\otimes_{i=1}^n \otimes_{j=1}^p P_\theta^{X_{[j]}^{(p)}} : \theta \in \Theta\}).$$

The upper index of the probability distributions is used to control the fixed set size of the ranked samples. Since we restrict ourselves to dominated experiments, we establish the existence or non-existence of Markov kernels between the experiments mentioned above (randomization criterion) and, by this, we decide on the existence of an informational order. The main relevant results are: The RSS experiment without repetition of size n is not more informative than the SRS experiment of size n. Therefore it exists at least a decision problem where a decision with the SRS gives more information than a decision with the RSS experiment without repetition of size n. Analogously, the RSS experiment with n repetitions is not more informative than the SRS of size np. The same implication for the information follows here. Another relevant result, perhaps not directly for the applications, but for the rest of the assertions, is that the RSS with n repetitions is more informative than the SRS of size n. Despite the intuitive thought that an observation of a RSS random variable would contain more information than a purely random observation, since for example it carries the additional information given by the prior visual ranking, we show that this is not true, since the experiment generated by a single RSS random variable

$$(\{\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^{X_{[i]}^{(n)}} : \theta \in \Theta\})$$

is not more informative than a purely random experiment

$$(\{\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^X : \theta \in \Theta\}).$$

In the Appendix of the thesis we relate RSS experiments with repetition to stratified random sampling experiments. The main result here is that the stratification generated by the RSS random variables is not the optimal stratification one can use for the estimation of the sample mean. The results are given for fixed but arbitrary sample sizes. All the results mentioned here hold under some regularity conditions for the respective families of probability distributions. The proofs rely basically on the sufficiency and completeness of the order statistic for specific families of probability distributions.

The Content of the Chapters

This dissertation is structured as follows: The first chapter is the Introduction. In the second chapter, the Prerequisites, we give the necessary notations and the basic results from the theory of experiments and order statistics. The purpose of the third chapter is to give a precise mathematical definition of the RSS random variables and of RSS statistical experiments. Starting from a new definition of RSS random variables via random stochastic matrices, we construct two models of RSS statistical experiments. The first model, the RSS experiments without repetition, consists of a family of experiments indexed by the number of RSS random variables to be considered in the model. The second model consists of the RSS experiments with repetition. They are defined as a product experiment of RSS experiments without repetition. The fourth chapter focuses on the comparisons in the family of RSS experiments without repetition. In the beginning of the chapter we recall and make some new relevant remarks on the order statistic as a sufficient, exhaustive and complete statistic. The main result of this chapter is that the RSS without repetition is not more informative than the SRS and that the family of RSS experiments without repetition is not more informative. In section 4.3 we treat some decision problems where RSS behaves more informative than the SRS. The content of the fifth chapter is the comparison of RSS experiments with repetition with a SRS of a smaller dimension. In this case, we prove the existence of a Markov kernel to assure the informational order, that RSS with repetition is more informative than the SRS. From an applied point of view, the result is of relative importance, even though we give two examples, to see how the existing Markov kernel generates a better decision. The second example shows that the comparison of an estimator based on n observations with an estimator based on a higher number of observations is in the case of subconvex loss functions motivated. In the next chapter we proceed with the comparison of the RSS experiment with repetition with the SRS of the same size and the result is that the RSS is not more informative than the SRS. In the last chapter, the Appendix, we treat the RSS technique from a different point of view. We introduce stratified random sampling as statistical experiments, the RSS falls into this group. Then, by using results by Taga [26], we conclude that the RSS experiment is not the optimal stratification one can choose for the estimation of the expectation. On the other hand, a reasonable compromise in this case, since the boundaries of the strata for the optimal stratification are computationally difficult to achieve.

Acknowledgments

I am especially grateful to Prof. Dr. M. Denker who supervised my Ph.D. thesis. I would like to thank him not only for the fruitful discussions during the preparation, but also for his kindness to teach me recognizing relevance. I thank him also for his trust in me and the freedom he gave me in doing research.

I would like to thank Prof. Dr. A. Munk for being the co-referee of my thesis.

Moreover, I would like to thank Dr. W. Schwarzwäller, Dr. M. Stadlbauer and Dr. G. Freitag for the great help they offered me during the preparation. Last but not least, I would like to express my appreciation to Julia and Burghardt, for being the wonderful friends and family in the last years and to all who supported me writing my thesis.

2 Prerequisites

In this chapter we repeat some definitions from statistical decision theory and recall some well-known facts, which are needed in the sequel.

2.1 Notations

If not otherwise specified, we will always denote by $\mathcal{B}$ the corresponding σ-Algebra of a measurable space E. All random variables are defined on a probability space $(\Omega, \mathcal{F}, P)$. Expectations with respect to P will be denoted by E_P. For a topological space D we denote by $\mathcal{B}(D)$ the Borel σ-Algebra and by $\mathcal{B}_0(D)$ the Baire σ-Algebra in D. The Borel σ-Algebra $\mathcal{B}(\mathbb{R}^n)$ is equal to the product σ-Algebra[3] $\mathcal{B}(\mathbb{R})^n$. If $T : (E, \mathcal{B}) \to (E', \mathcal{B}')$ is a measurable map then we denote by P^T the image measure on $\mathcal{B}'$. For the probability space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, P)$ we write $P(B)$ for an arbitrary set $B \in \mathcal{B}(\mathbb{R})^n$ and $P(x)$ for the distribution function. If $\mathbb{T}$ is a matrix then we denote by T_{i*} (respectively T_{*j}) the i-th row vector (respectively the j-th column vector) of the matrix. We denote by $\mathcal{M}(\mathcal{B}(\mathbb{R})^n)$ the set of all probability measures on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$.

2.2 Comparison of Experiments

Definition 1 (Statistical experiment) Let $\Theta \neq \emptyset$ be an arbitrary set. A statistical experiment for the parameter space Θ is a triple $\mathcal{E} = (E, \mathcal{B}, \mathcal{P})$ where $(E, \mathcal{B})$ is a measurable space called the sample space and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is a family of probability measures on $\mathcal{B}$. The collection of all experiments for the parameter space Θ is denoted by $\mathcal{E}(\Theta)$.

Remark 2 We say that a statistical experiment $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ is dominated if it exists a σ-finite measure ν on $\mathcal{B}$ such that $P_\theta \ll \nu$, for every $\theta \in \Theta$.

Remark 3 In a parametric setting we could consider Θ to be an open subset in $\mathbb{R}^k$, $k \geq 1$. In a nonparametric setting we do not make any assumptions on the distribution of the random variables in the sample space. In this case the parameter space can be taken to be a function space, for example the space of all continuous distribution functions on the sample space.

Definition 4 (Markov Kernel) Let $(E, \mathcal{B})$ and $(E', \mathcal{B}')$ be two measurable spaces. A kernel from $(E, \mathcal{B})$ to $(E', \mathcal{B}')$ is a function on $E \times \mathcal{B}'$ with the properties:

- $x \mapsto K(x, B')$ is $\mathcal{B}$ measurable for all $B' \in \mathcal{B}'$
- $B' \to K(x, B')$ is a measure on $\mathcal{B}'$ for all $x \in E$.

The kernel K is a Markov kernel if $K(x, E') = 1$ for all $x \in E$.

[3] See BAUER [3], section 6.

Let $\mathcal{E} = (E, \mathcal{B}, \mathcal{P})$ be a statistical experiment and let D be a topological space. We call $(D, \mathcal{B}_0(D))$ the decision space. Note that if D is a metric space then $\mathcal{B}_0(D) = \mathcal{B}(D)$.

Definition 5 (Decision Function) Let $(D, \mathcal{B}_0(D))$ be a decision space. Any Markov kernel $\rho : E \times \mathcal{B}_0(D) \to [0, 1]$ is called a *decision function* and the set of all decision functions for an experiment is denoted by $\mathcal{R}(\mathcal{E}, D)$. A decision function $\rho \in \mathcal{R}(\mathcal{E}, D)$ is said to be *nonrandomized* if $\rho(\cdot, B) \in \{0, 1\}$, $P_\theta - a.s$ for every $\theta \in \Theta$ and $B \in \mathcal{B}_0(D)$. The set of all nonrandomized decision functions will be denoted by $\mathcal{R}_0(\mathcal{E}, D)$.

Definition 6 (Loss Function) A family $(W_\theta)_{\theta\in\Theta}$ of functions $W_\theta : D \to \mathbb{R}$, $\theta \in \Theta$, $\mathcal{B}_0(D)$ measurable, is called a *loss function*. We say that $(W_\theta)_{\theta\in\Theta}$ is a lower semicontinuous loss function if each W_θ is bounded from below ($\inf_t W_\theta(t) > -\infty$) and the sets $\{W_\theta(t) \leq a\}$, $a \in \mathbb{R}$ are closed. We say $(W_\theta)_{\theta\in\Theta}$ is a continuous loss functions if each W_θ is continuous.

Definition 7 (Risk Function) The *risk* of the decision $\rho \in \mathcal{R}(\mathcal{E}, D)$ at $\theta \in \Theta$ is denoted by

$$W_\theta \rho P_\theta := \int_E \int_D W_\theta(t) \rho(x, dt) P_\theta(dx),$$

and the function $\theta \mapsto W_\theta \rho P_\theta$, $\theta \in \Theta$ is called the *risk function* of the decision ρ.

Remark 8 If $\rho \in \mathcal{R}_0(\mathcal{E}, D)$ then, obviously,

$$W_\theta \rho P_\theta = \int_E W_\theta(\rho(x)) P_\theta(dx), \quad \theta \in \Theta.$$

Definition 9 (Decision Problem) The triple (Θ, D, W) consisting of a parameter space Θ, a decision space D and a loss function $(W_\theta)_{\theta\in\Theta}$ is called a decision problem (for Θ).

Remark 10 Here, Θ can be identified with a fixed family of probability distributions.

In what follows we extend the definitions usually used in common decision theory to definitions used in the theory of the comparison of statistical experiments. For further details see LeCam [19], Strasser [25] or Shiryaev-Spokoiny [23]. Let $(E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ be a statistical experiment. We denote by $\mathcal{C}_b(D)$ the set of all real continuous and bounded functions on D, $\|f\| := \sup_{t\in D} |f(t)|$ and by $L(\mathcal{E}) := \{\mu \in ca(E, \mathcal{B}) : \sigma \perp P_\theta, \theta \in \Theta \Rightarrow \sigma \perp \mu\}$ where $ca(E, \mathcal{B})$ is the set of all bounded, signed measures on $(E, \mathcal{B})$, $\|\mu\| := (\mu^+ + \mu^-)(E)$. For convenience we denote

$$f\rho\mu = \int \int f(t) \rho(x, dt) \mu(dx)$$

if $f \in \mathcal{C}_b(D)$ and $\mu \in L(\mathcal{E})$. Every decision function $\rho \in \mathcal{R}(\mathcal{E}, D)$ defines a bilinear function $\beta_\rho : \mathcal{C}_b(D) \times L(\mathcal{E}) \to \mathbb{R}$ according to

$$\beta_\rho(f, \mu) := f\rho\mu, \quad f \in \mathcal{C}_b(D), \mu \in L(\mathcal{E}).$$

Definition 11 (Generalized Decision Function) A *generalized decision function* for the statistical experiment $\mathcal{E} = (E, \mathcal{B}, \mathcal{P})$ and D is a bilinear function $\beta : \mathcal{C}_b(D) \times L(\mathcal{E}) \to \mathbb{R}$ satisfying the following conditions:

1. $|\beta(f, P)| \leq \|f\|_\infty \|P\|$, if $f \in \mathcal{C}_b(D)$, $P \in L(\mathcal{E})$.
2. $\beta(f, P) \geq 0$, if $f \geq 0$, $P \geq 0$.
3. $\beta(1, P) = P(E)$, if $P \in L(\mathcal{E})$.

The set of all generalized decision functions is denoted by $\mathcal{B}(\mathcal{E}, D)$. For every $\rho \in \mathcal{R}(\mathcal{E}, D)$ we have $\beta_\rho \in \mathcal{B}(\mathcal{E}, D)$.

Definition 12 (More Informative Experiments) Let $\mathcal{E}_1 = (E_1, \mathcal{B}_1, \{P_\theta : \theta \in \Theta\})$ and $\mathcal{E}_2 = (E_2, \mathcal{B}_2, \{Q_\theta : \theta \in \Theta\})$ be two statistical experiments and (Θ, D, W) be a decision problem such that $(W_\theta)_{\theta\in\Theta}$ is a lower semicontinuous loss function. The experiment $\mathcal{E}_1$ is called *more informative* than the experiment $\mathcal{E}_2$ *for the decision problem* (Θ, D, W) denoted by $\mathcal{E}_2 \overset{(D,W)}{\subseteq} \mathcal{E}_1$, if for every generalized decision function $\beta_2 \in \mathcal{B}(\mathcal{E}_2, D)$ there exists $\beta_1 \in \mathcal{B}(\mathcal{E}_1, D)$ such that

$$\beta_1(W_\theta, P_\theta) \leq \beta_2(W_\theta, Q_\theta), \quad \theta \in \Theta. \tag{1}$$

If $\mathcal{E}_2 \overset{(D,W)}{\subseteq} \mathcal{E}_1$ holds for every (D, W) with a continuous and bounded loss function, then we denote $\mathcal{E}_2 \subseteq \mathcal{E}_1$. The relation " $\subseteq$ " is an order relation on the space of experiments $\mathcal{E}(\Theta)$. If $\mathcal{E}_2 \subseteq \mathcal{E}_1$ then $\mathcal{E}_1$ is called *more informative* than $\mathcal{E}_2$ for the parameter space Θ. If neither $\mathcal{E}_2 \subseteq \mathcal{E}_1$ nor $\mathcal{E}_1 \subseteq \mathcal{E}_2$, then we say that the experiments $\mathcal{E}_1$ and $\mathcal{E}_2$ are *not comparable* for the parameter space Θ. We say that a family of statistical experiments $\mathcal{E}_{\{i\in I\}}$, with an ordered index set I, is more informative if $\mathcal{E}_{i+1}$ is more informative than $\mathcal{E}_i$ for every $\{i, i+1\} \in I$.

Remark 13 Even though according to Definition 9, a decision problem is given by the triple (Θ, D, W), in the notation for the informational order, the parameter space for which the comparison between the experiments is done, is omitted and assumed to be fixed.

Remark 14 Let $\mathcal{E}_1$ and $\mathcal{E}_2$ be dominated experiments and let (Θ, D, W) be a decision problem such that D is a locally compact space with countable base and $(W_\theta)_{\theta\in\Theta}$ is a continuous loss function. The experiment $\mathcal{E}_1$ is *more informative* than the experiment $\mathcal{E}_2$ for the decision problem (Θ, D, W), i.e. $\mathcal{E}_2 \overset{(D,W)}{\subseteq} \mathcal{E}_1$, if for every $\rho_2 \in \mathcal{R}(\mathcal{E}_2, D)$ there is $\rho_1 \in \mathcal{R}(\mathcal{E}_1, D)$ such that

$$W_\theta \rho_1 P_\theta \leq W_\theta \rho_2 Q_\theta, \quad \theta \in \Theta. \tag{2}$$

For a proof, see Strasser [25], Section 43.

The following definition of the randomization of experiments is given for the particular case of dominated experiments on locally compact spaces, for the general definition see Strasser [25], Section 55.

Definition 15 (Randomization of Experiments) Suppose $\mathcal{E}_1 = (E_1, \mathcal{B}_1, \{P_\theta : \theta \in \Theta\})$ is a dominated experiment and $\mathcal{E}_2 = (E_2, \mathcal{B}_2, \{Q_\theta : \theta \in \Theta\})$ is such that E_2 is a locally compact space with countable base and $\mathcal{B}_2 = \mathcal{B}(E_2)$. Then $\mathcal{E}_2$ is a *randomization* of $\mathcal{E}_1$ if there is a Markov kernel K from $(E_1, \mathcal{B}_1)$ to $(E_2, \mathcal{B}_2)$ such that

$$Q_\theta(A) = \int K(x, A) P_\theta(dx)$$

for every $\theta \in \Theta$ and $A \in \mathcal{B}_2$.

Analogously, the following theorem has a more general context, for the general randomization case, stipulated first by LeCam [19]. This version is in the spirit of Heyer [16].

Theorem 16 Let $\mathcal{E}_1 = (E_1, \mathcal{B}_1, \{P_\theta : \theta \in \Theta\})$ be a dominated statistical experiment and $\mathcal{E}_2 = (E_2, \mathcal{B}_2, \{Q_\theta : \theta \in \Theta\})$ be a statistical experiment such that E_2 is locally compact with countable base and $\mathcal{B}_2 = \mathcal{B}(E_2)$. Then $\mathcal{E}_2 \subseteq \mathcal{E}_1$ (i.e. $\mathcal{E}_1$ is more informative then $\mathcal{E}_2$) if and only if $\mathcal{E}_2$ is a randomization of $\mathcal{E}_1$.

Proof.
See for example Theorem 55.9 and Corollary 55.11 in Strasser [25]. □

Sufficiency, Exhaustivity

Definition 17 (Conditional Probability Distribution) Let $X : (\Omega, \mathcal{F}, P) \to (E_1, \mathcal{B}_1)$ be a random variable and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-subfield. The conditional distribution of X given $\mathcal{G}$ is any Markov kernel $P^{X|\mathcal{G}}$ from $(\Omega, \mathcal{G})$ to $(E_1, \mathcal{B}_1)$ such that for all $B_1 \in \mathcal{B}_1$, $\omega \mapsto P^{X|\mathcal{G}}(\omega, B_1)$ is a version of the conditional probability $P(X^{-1}(B_1) \mid \mathcal{G})$.

Proposition 18 Let $X : (\Omega, \mathcal{F}, P) \to (E_1, \mathcal{B}_1)$ be a random variable and E_1 be a Polish space. Then for every σ-subfield $\mathcal{G} \subseteq \mathcal{F}$ there exists the conditional distribution $P^{X|\mathcal{G}}$. Two versions of the conditional probability are equal up to a set of measure 0.

Proof.
See Proposition 56.5 in Bauer [3]. □

Definition 19 (Sufficiency) Let $(E_1, \mathcal{B}_1)$ be a measurable space and $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a set of probability measures on $\mathcal{B}_1$. A σ-Algebra $\mathcal{B}_0 \subseteq \mathcal{B}_1$ is said to be $\mathcal{P}$-sufficient if for every $B \in \mathcal{B}_1$ there is a $\mathcal{B}_0$-measurable function f_B such that $f_B = P_\theta(B \mid \mathcal{B}_0)\ P_\theta - a.e.$ for every $\theta \in \Theta$. A measurable map $S : (E_1, \mathcal{B}_1) \to (E_2, \mathcal{B}_2)$ is called a *sufficient statistic*[4] if $S^{-1}\mathcal{B}_2$ is $\mathcal{P}$-sufficient. Then we also write $f_B(x) = P(B \mid S(x))$, $x \in E_1$.

[4]Historical Remark: Fisher developed the idea that a statistic S is sufficient if any statistic T has a conditional distribution, given S, which is independent of the probability measure. Knowing $T(x)$, in addition to $S(x)$, can therefore, contribute nothing to the knowledge about the 'true' probability measure.

Remark 20 (1) By Proposition 18, if E_1 is a Polish space and $\mathcal{B}_0 \subseteq \mathcal{B}_1$ a σ-subfield, then there exist Markov kernels $K_\theta : E_1 \times \mathcal{B}_1 \to [0,1]$ such that

$$K_\theta(x, B) := P_\theta(B \mid \mathcal{B}_0)(x) \quad P_\theta - a.s., \quad B \in \mathcal{B}_1$$

If $\mathcal{B}_0$ is sufficient, then we can choose K_θ independent of θ and the measurable functions f_B in Definition (19) can be determined by the Markov kernels.
In this case, by definition of the conditional probability, the Markov kernels are $\mathcal{B}_0$ measurable solutions of the equation:

$$\int_C K(x, B)(P_\theta \mid_{\mathcal{B}_0})(dx) = \int_C 1_B(x) P_\theta(dx) = P_\theta(B \cap C), \quad B \in \mathcal{B}_1$$

for all $C \in \mathcal{B}_0$ and $\theta \in \Theta$.
(2) Analogously, if E_1 is a Polish space and if $S : (E_1, \mathcal{B}_1) \to (E_2, \mathcal{B}_2)$ is a sufficient statistic then it exists the Markov kernel $Q : E_2 \times \mathcal{B}_1 \to [0,1]$,

$$Q(s, B) := P_\theta(B \mid S = s) \quad P_\theta^S - a.s., \quad B \in \mathcal{B}_1$$

such that the right hand side does not depend P_θ^S-a.s. on $\theta \in \Theta$.
In this case, by the definition of factorised conditional probability, the equality holds

$$\int_{C_2} Q(s, B) P_\theta^S(ds) = \int_{S^{-1}(C_2)} 1_B(x) P_\theta(dx) = P_\theta(B \cap S^{-1}(C_2))$$

for every $C_2 \in E_2, \quad B \in \mathcal{B}_1$ and $\theta \in \Theta$.

Definition 21 (Minimal Sufficiency) The sufficient statistic $S_* : (E_1, \mathcal{B}_1) \to (E_3, \mathcal{B}_3)$ is said to be *minimal sufficient* if for any sufficient statistic $S : (E_1, \mathcal{B}_1) \to (E_2, \mathcal{B}_2)$ there exists a function $H : (E_2, \mathcal{B}_2) \to (E_3, \mathcal{B}_3)$ such that $S_* = H \circ S$, $P_\theta - a.s.$, for every $\theta \in \Theta$.

Definition 22 (Exhaustivity) Let $\mathcal{E}_i = (E_i, \mathcal{B}_i, \{P_{i,\theta} : \theta \in \Theta\})$, $i = 1, 2$, be statistical experiments. $\mathcal{E}_2$ is called *exhaustive*[5] for $\mathcal{E}_1$ if there is a Markov kernel K from $(E_2, \mathcal{B}_2)$ to $(E_1, \mathcal{B}_1)$ such that

$$P_{1,\theta}(A_1) = \int K(x, A_1) P_{2,\theta}(dx) \tag{3}$$

for every $\theta \in \Theta$ and every $A_1 \in \mathcal{B}_1$.
We call $S : (E_1, \mathcal{B}_1) \to (E_2, \mathcal{B}_2)$ an *exhaustive statistic* for the experiment $\mathcal{E}_1$ if there is a Markov kernel Q from $(E_2, \mathcal{B}_2)$ to $(E_1, \mathcal{B}_1)$ such that

$$P_{1,\theta}(A_1) = \int Q(y, A_1) P_{1,\theta}^S(dy) \tag{4}$$

for every $\theta \in \Theta$ and every $A_1 \in \mathcal{B}_1$.

[5]The concept of *exhaustivity* goes back to Blackwell [5]

Example
Let $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ be a statistical experiment and $B_0 \subseteq \mathcal{B}$ a σ-subfield. Denote by $\mathcal{E} \mid B_0 := (E, \mathcal{B}_0, \{P_\theta |_{\mathcal{B}_0} : \theta \in \Theta\})$ the statistical experiment restricted to the σ-subfield $\mathcal{B}_0$. Then $\mathcal{E}$ is exhaustive for $\mathcal{E} \mid B_0$. To prove this note that $K(x, A) = 1_A(x)$ for $A \in \mathcal{B}_0$ and $x \in E$ defines a Markov kernel $K : E \times \mathcal{B}_0 \to [0, 1]$ from $\mathcal{E}$ to $\mathcal{E} \mid B_0$.

Remark 23 (Interpretation of Exhaustivity) By Theorem 1.10.33 in Pfanzagl [21], if $(E_1, \mathcal{B}_1)$ is a Polish space, then for every Markov kernel K from $(E_2, \mathcal{B}_2)$ to $(E_1, \mathcal{B}_1)$ there exists a measurable function $m : E_2 \times (0, 1) \to E_1$ such that for every $y \in E_2$, the probability measure $K(y, \cdot)$ is the image of the Lebesgue measure $\mathcal{U}$ on $(0, 1)$ by the transformation $u \mapsto m(y, u)$ i.e.

$$K(y, B) = \mathcal{U}\{u \in (0, 1) : m(y, u) \in B\}, \quad B \in \mathcal{B}_2.$$

Therefore, if $\mathcal{E}_2$ is exhaustive for $\mathcal{E}_1$, i.e. there exists a Markov kernel such that

$$P_{1,\theta}(A_1) = \int K(y, A_1) P_{2,\theta}(dy), \quad \theta \in \Theta.$$

Thus, we obtain

$$P_{2,\theta} \otimes \mathcal{U}\{m(y, u) \in A_1\} = P_{1,\theta}(A_1)$$

for every $\theta \in \Theta$ and $A_1 \in \mathcal{B}_1$. If we know y and determine a realization u from the uniform distribution over $(0, 1)$, $m(y, u)$ defines a random variable with exactly the same distribution as the original one, for every $\theta \in \Theta$.

Proposition 24

1. A sufficient statistic is exhaustive if $(E_1, \mathcal{B}_1)$ is a Polish space.
2. Any exhaustive statistic is sufficient.

Proof.
(1) Obvious, by remark (20).
(2) See Theorem 1.3.9. in Pfanzagl [21]. □

Remark 25 The fact that any exhaustive statistic is sufficient does, however, not involve that the Markov kernel which occurs in the definition of exhaustivity is a conditional distribution given the sufficient statistic. The situation is different if the sufficient statistic is minimal.

Proposition 26 Assume that $\mathcal{B}_1$ is countably generated and $S : (E_1, \mathcal{B}_1) \to (E_2, \mathcal{B}_2)$ is a minimal sufficient statistic. Then any Markov kernel Q from $(E_2, \mathcal{B}_2)$ to $(E_1, \mathcal{B}_1)$ fulfilling

$$P_\theta(A) = \int Q(y, A) P_\theta^S(dy) \tag{5}$$

for every $\theta \in \Theta$ and $A \in \mathcal{B}_1$ is a conditional distribution $P_\theta(\cdot \mid S = s)$ and therefore unique in the following sense: if Q_i, $i = 1, 2$ fulfill (5), then $Q_1(y, \cdot) = Q_2(y, \cdot)$, $P_\theta^S - a.s.$.

Proof.
See Proposition 1.4.9. in Pfanzagl [21]. $\square$

Lemma 27 Suppose that $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ is an experiment and $F : (E, \mathcal{B}) \to (E_1, \mathcal{B}_1)$ is a measurable mapping. Then the experiments $F_*\mathcal{E} := (E_1, \mathcal{B}_1, \{P_\theta \circ F^{-1} : \theta \in \Theta\})$ and $\mathcal{E} |_{F^{-1}\mathcal{B}_1} := (E, F^{-1}\mathcal{B}_1, \{P_\theta |_{F^{-1}\mathcal{B}_1} : \theta \in \Theta\})$ are mutually exhaustive.

Proof.
Define the kernel $K_1 : E \times \mathcal{B}_1 \to [0, 1]$,

$$K_1(x, A) = (1_A \circ F)(x), \quad x \in E,\, A \in \mathcal{B}_1$$

Then we have that

$$P_\theta \circ F^{-1}(A) = \int K_1(x, A)\, dP_\theta(x) \quad \forall \theta \in \Theta,\, A \in \mathcal{B}_1.$$

Therefore $\mathcal{E}$ is exhaustive for $F_*\mathcal{E}$. It is clear that $F_*\mathcal{E} = F_*(\mathcal{E} |_{F^{-1}\mathcal{B}_1})$. This implies that $\mathcal{E} |_{F^{-1}\mathcal{B}_1}$ is exhaustive for $F_*\mathcal{E}$. To prove the converse define $K_2 : \mathcal{E}_1 \times F^{-1}\mathcal{B}_1 \to [0, 1]$,

$$K_2(x_1, A) = 1_{F(A)}(x_1), \quad A \in F^{-1}\mathcal{B}_1,\, x_1 \in E_1.$$

Note that $A = F^{-1}(A_1)$ implies $F^{-1}(F(A)) = F^{-1}(A_1)$. It follows that

$$\begin{aligned} P_\theta(A) &= P_\theta(F^{-1}(A_1)) = \int 1_{F^{-1}(A_1)}(x)\, dP_\theta(x) \\ &= \int 1_{F^{-1}(F(A))}(x)\, dP_\theta(x) = \int 1_{F(A)}(x_1)\, dP_\theta \circ F^{-1}(x_1). \end{aligned}$$

$\square$

Definition 28 (Completeness) A family of probability distributions $\{P : P \in \mathcal{P}\}$ on $(E, \mathcal{B})$ is q-complete, $1 \le q \le \infty$, if for every measurable $f \in \bigcap_{P \in \mathcal{P}} L^q(P)$,

$$\int f(x)P(dx) = 0\, \forall P \in \mathcal{P} \;\Rightarrow\; f = 0\, P - a.s.\ \forall P \in \mathcal{P}.$$

We call $S : (E, \mathcal{B}) \to (E_1, \mathcal{B}_1)$ a q-complete statistic if $\{P^S : P \in \mathcal{P}\}$ is a q-complete family of probability measures. When we say a family of probability distributions is complete, we mean 1-completeness.

Definition 29 (Symmetrically Completeness) A family of probability distributions $\{P : P \in \mathcal{P}\}$ on $(E, \mathcal{B})$ is *symmetrically q-complete of order* n if for any real permutation invariant function f_n on E^n, with $f_n \in \bigcap_{P \in \mathcal{P}} \text{Ł}^q(P^n)$ we have

$$\int f_n(x_1, \ldots, x_n) \prod_{i=1}^n P(dx_i) = 0\, \forall P \in \mathcal{P} \;\Rightarrow\; f_n = 0\, P^n - a.s.\ \forall P \in \mathcal{P}.$$

Proposition 30 Every sufficient and complete statistic on a Polish space is minimal sufficient.

Proof.
See Proposition 1.4.8 in Pfanzagl [21]. □

Lemma 31 (Plachky-Landers-Rogge) For $i \in \{1, \ldots, n\}$ let $(\mathcal{E}_i, \mathcal{B}_i)$ be measurable spaces and $\mathcal{P}_i$ q-complete families of probability measure on $\mathcal{B}_i$. Then the family

$$\{\otimes_{i=1}^n P_i : \; P_i \in \mathcal{P}_i, \; i = 1, ..., n\}$$

is q-complete.

Proof.
See Proposition 1.5.6 in Pfanzagl [21]. □

Proposition 32 (Mandelbaum-Rüschendorf) Assume that $\mathcal{P}$ is q-complete and closed under convex combinations. Then $\mathcal{P}$ is symmetrically q-complete of order n for every $n \in \mathbb{N}$.

Proof.
See for example Proposition 7.78 in Witting [29]. □

2.3 Order Statistics

Let $n \in \mathbb{N}$ be fixed and arbitrary. Denote the group of permutations on $\{1, ..., n\}$ by $\mathcal{S}_n$. Define the map $\tau_\pi : \mathbb{R}^n \to \mathbb{R}^n$ by $\tau_\pi((x_1, \ldots, x_n)) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every $\pi \in \mathcal{S}_n$.

Definition 33 (Order Statistic) The map $O_n : \mathbb{R}^n \to \mathbb{R}^n$ defined as

$$O_n((x_1, \ldots, x_n)) = (y_1, \ldots, y_n)$$

with $y_1 \leq \ldots \leq y_n$ and it exists a permutation $\pi \in \mathcal{S}_n$ such that $\tau_\pi(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ is called *order statistic*.

Let $X, X_i : (\Omega, \mathcal{F}, P) \to \mathbb{R}$, $1 \leq i \leq n$ be a sample of i.i.d. random variables each having probability distribution P^X. The order statistic for the sample of size n is given by

$$(X_{(1:n)}, \ldots, X_{(n:n)}) := O_n((X_1, \ldots, X_n))$$

and the probability distribution by $P^{O_n X^n}$ where $X^n := (X_1, \ldots, X_n)$. The i-th order statistic in a sample of size n is given by

$$X_{(i:n)} := \mathrm{pr}_{(i:n)} O_n(X_1, \ldots, X_n), \quad 1 \leq i \leq n$$

and we denote the probability distribution by $P^{X_{(i:n)}}$. The probability distribution of two order statistics is denoted by $P^{X_{(i:n)},X_{(j:n)}}$.

The Exact Probability Distribution of the Order Statistic

Proposition 34 Let X_i, $1 \leq i \leq n$ be i.i.d. real random variables each with distribution P^X. For each $1 \leq i < j \leq n$ and $y, z \in \mathbb{R}$ and by denoting $\pi_{nl}(P^X(z)) := \binom{n}{l}(P^X(z))^l(1 - P^X(z))^{n-l}$ and

$$\pi_{n,m,l}(P^X(y), P^X(z)) := \frac{n!}{m!l!(n-m-l)!}(P^X(y))^m(P^X(z) - P^X(y))^l(1 - P^X(z))^{n-l-m}$$

we have

$$\begin{aligned} P^{X_{(i:n)}}(z) &= \sum_{l=1}^{n} \pi_{nl}(P^X(z)), \\ P^{X_{(i:n)},X_{(j:n)}}(y, z) &= \begin{cases} P(X_{(j:n)} \leq z) & z \leq y \\ \sum_{m=i}^{n} \sum_{l=0 \vee (j-m)}^{n-m} \pi_{n,m,l}(P^X(y), P^X(z)) & y < z \end{cases} \end{aligned}$$

Proof.
See for example Witting [29], section 7.2.2. □

Denote for $1 \leq i \leq n$ the density with respect to to the Lebesgue measure of the Beta distribution, $\beta_{(i:n)} : [0, 1] \to \mathbb{R}$,

$$\beta_{(i:n)}(u) = n! \frac{u^{i-1}}{(i-1)!} \frac{(1-u)^{n-i}}{(n-i)!}, \quad 1 \leq i \leq n. \tag{6}$$

Analogously, denote the density with respect to the Lebesgue measure of a m-dimensional Beta distribution by

$$\beta_{(i_1,\ldots,i_m:n)}(u_1, \ldots, u_m) = n! \frac{u^{i_1-1}}{(i_1-1)!} \frac{(u_2-u_1)^{i_2-i_1-1}}{(i_2-i_1-1)!} \cdots \frac{(1-u_m)^{n-i_m}}{(n-i_m)!} \tag{7}$$

for $0 \leq u_1 < \ldots < u_m \leq 1$ and 0 else.

Proposition 35 Let X_i, $1 \leq i \leq n$ be i.i.d. real random variables each with the continuous distribution P^X. Then the following holds:

1. $P^{X_{(i:n)}} \ll P^X$ and
$$\frac{dP^{X_{(i:n)}}}{dP^X}(x) = \beta_{(i:n)}(P^X(x)), \quad P^X - a.s. \tag{8}$$

2. Generally, for $1 \leq i_1 < \ldots < i_m \leq n$
$$\frac{dP^{X_{(i_1:n)},\ldots,X_{(i_m:n)}}}{d(P^X)^m}(x_1, \ldots, x_m) = \beta_{(i_1,\ldots,i_m:n)}(P^X(x_1), \ldots, P^X(x_m)), \; (P^X)^m - a.s. \tag{9}$$

Proof.
See for example Witting [29], section 7.2.2. □

3 Ranked Set Sampling Experiments

In this chapter we define the Ranked Set Sampling (RSS) random variables to generate the RSS statistical experiments by making use of the properties of random stochastic matrices.

3.1 Sampling Random Variables

Consider $n \in \mathbb{N}$ to be fixed but arbitrary and $1 \le r \le n$. A random matrix $(T_{ij})_{1\le i,j\le n}$: $(\Omega, \mathcal{F}, P) \to \mathbb{R}^{n\times n}$ is called a *random stochastic matrix* if $\sum_{j=1}^{n} T_{ij} = 1$ holds $P - a.s.$ for every $1 \le i \le n$. If the condition $\sum_{i=1}^{n} T_{ij} = 1$ also holds $P - a.s.$ for every $1 \le j \le n$, then the matrix is called a *random double stochastic matrix*.We call a random stochastic matrix $\mathbb{T} \in \mathbb{R}^{n\times n}$ a *sampling matrix of size* r if the following conditions are fulfilled:

1. $T_{ij} \in \{0, 1\}$, $1 \le i, j \le n$
2. $T_{ij} = 0$, P-a.s., $r + 1 \le j \le n$, $1 \le i \le r$
3. $T_{ii} = 1$, P-a.s., $r + 1 \le i \le n$
4. $\mathbb{T}_{i_1*}$ independent of $\mathbb{T}_{i_2*}$ for all $1 \le i_1 \ne i_2 \le n$.

Henceforth we denote such a random matrix by $\mathbb{T}^{(r)} := (T_{ij}^{(r)})_{1\le i,j\le n}$ where the upper index corresponds to the size of the sampling matrix.

Remark 36 The rows of the sampling matrix $\mathbb{T}_{i*}^{(r)}$, $1 \le i \le n$, have a multinomial distribution,

$$\mathbb{T}_{i*}^{(r)} \sim Mn(1, p_{i1}, \dots, p_{in}), \quad \sum_{j=1}^{n} p_{ij} = 1,\ p_{ij} = 0\ \ \forall j \ge r+1$$

where we have denoted by $p_{ij} := P(T_{ij}^{(r)} = 1)$, $1 \le i, j \le n$. For each $t := (t_j)_{1\le j\le n}$, such that $t_j \in \{0, 1\}$ and $\sum_{j=1}^{n} t_j = 1$ we have $P(T_{i1}^{(r)} = t_1, \dots, T_{in}^{(r)} = t_n) = \prod_{j=1}^{r} (p_{ij})^{t_j}$.

Let $X : (\Omega, \mathcal{F}, P) \to \mathbb{R}$ be a random variable with distribution P^X. Denote by $\mathbb{X} := (X_{ij})_{1\le i,j\le n}$ a matrix of i.i.d. real random variables each with distribution P^X.

Definition 37 (Sampling Random Variables) Let $\mathbb{T}^{(r)}$ be a sampling matrix of size r and let $\mathbb{X}$ be a matrix of i.i.d. real random variables each with distribution P^X, such that $\mathbb{X}_{i*}$ is independent of $(\mathbb{T}_{k*}^{(r)})_{1\le k\ne i\le n}$. Then the *sampling random variables of size* r are given by

$$X_{[i]}^{(r)} := (\mathbb{T}_{i*}^{(r)})' \mathbb{X}_{i*}, \quad 1 \le i \le n. \tag{10}$$

Lemma 38 The distribution of the sampling random variables of size r satisfies:

1. $X_{[i]}^{(r)} \sim P^{X_{[i]}^{(r)}}$, where

$$P^{X_{[i]}^{(r)}}(A) = \begin{cases} \sum_{j=1}^{r} P(X_{ij} \in A,\, T_{ij}^{(r)} = 1), & 1 \leq i \leq r \\ P^X(A), & r+1 \leq i \leq n \end{cases}$$

for every $A \in \mathbb{R}$.

2. The Fundamental Equation

$$P^X(A) = \frac{1}{n}\left(\sum_{i=1}^{r} P^{X_{[i]}^{(r)}}(A) + (n-r)P^X(A)\right), \quad A \in \mathcal{B}(\mathbb{R}). \tag{11}$$

3. The sampling random variables of size r are independent random variables.

Proof.
Let $A \in \mathcal{B}(\mathbb{R})$.
1. If $1 \leq i \leq r$, then

$$\begin{aligned} P(X_{[i]}^{(r)} \in A) &= P(\sum_{j=1}^{r} T_{ij}^{(r)} X_{ij} \in A) \\ &= P(\left\{\sum_{j=1}^{r} T_{ij}^{(r)} X_{ij} \in A\right\} \cap \{\bigcup_{j=1}^{r}\{T_{ij}^{(r)} = 1\}\}) \\ &= \sum_{j=1}^{r} P(X_{ij} \in A,\, T_{ij}^{(r)} = 1). \end{aligned}$$

If $r+1 \leq i \leq n$ the assertion is obvious.

2. Let $\mathbb{F}^1, \ldots, \mathbb{F}^r$ be i.i.d. random double stochastic matrices such that $F_{ij}^k \in \{0,1\}$ for all $1 \leq i,j \leq n$ and $1 \leq k \leq r$. We assume $\mathbb{F}_{i*}^i = \mathbb{T}_{i*}^{(r)}$ -P.a.s for all $1 \leq i \leq r$. Then it follows that

$$\begin{aligned} &\sum_{i=1}^{n} P^{X_{[i]}^{(r)}}(A) = \sum_{i=1}^{r} P^{X_{[i]}^{(r)}}(A) + (n-r)P^X(A) \\ = &\sum_{i=1}^{r}\sum_{j=1}^{r} P(X_{ij} \in A, T_{ij}^{(r)} = 1) + (n-r)P^X(A) \\ = &\sum_{i=1}^{r}\sum_{j=1}^{r} P(X_{ij} \in A, F_{ij}^{i} = 1) + (n-r)P^X(A) \\ = &\sum_{i=1}^{r}\sum_{j=1}^{r} P(X_{1j} \in A, F_{ij}^{1} = 1) + (n-r)P^X(A) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{r}\sum_{j=1}^{r}\int_{\{F^1_{ij}=1\}} 1_{\{X_{1j}\in A\}}dP + (n-r)P^X(A) \\
&= \sum_{j=1}^{r}\int_{\{\bigcup_{i=1}^{r} F^1_{ij}=1\}} 1_{\{X_{1j}\in A\}}dP + (n-r)P^X(A) \\
&= \sum_{j=1}^{r}\int 1_{\{X_{1j}\in A\}}dP + (n-r)P^X(A) \\
&= nP^X(A)
\end{aligned}
$$

3. The assertion is obvious, since the rows of the sampling matrix of size r, $\mathbb{T}^{(r)}_{i*}$ are independent and $\mathbb{X}_{i*}$ is independent of $\mathbb{X}_{k*}$ and $\mathbb{T}^{(r)}_{k*}$ for all $1 \le k \ne i \le n$.

3.2 Ranked Set Sampling Random Variables

Let $\mathcal{M}(\mathcal{B}(\mathbb{R})^2)$ be the set of all probability distributions on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$. Denote by $\Theta_0 := \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{continuous}\}$ and let $(X_{ij}, Y_{ij})_{1\le i,j\le n} : (\Omega, \mathcal{F}, P) \to \mathbb{R}^2$ be a random matrix of i.i.d. random vectors, $(X_{ij}, Y_{ij}) \sim P^{XY}_\theta$, $\theta \in \Theta_0$, $1 \le i, j \le n$. Denote $\mathbb{X} := (X_{ij})_{1\le i,j\le n}$ (respectively $\mathbb{Y} := (Y_{ij})_{1\le i,j\le n}$) with the marginal distribution $X_{ij} \sim P^X_\theta$, $1 \le i, j \le n$ and independent (respectively $Y_{ij} \sim P^Y_\theta$, $1 \le i, j \le n$ and independent).

Remark 39 For every $1 \le i \le n$ let B_i be the ω-set, where $Y_{ij}(\omega) = Y_{ik}(\omega)$ for some distinct pair of integers j, k, then by continuity $P(B_i) = 0$. For every $1 \le i \le n$ we remove B_i from the space Ω on which $\mathbb{Y}$ is defined, and this will leave the joint distribution unchanged and make ties impossible.

Proposition 40 Every random double stochastic matrix of dimension n with $0-1$ entries is uniquely determined by a random permutation defined on the group $\mathcal{S}_n$ of permutations on $\{1, \dots, n\}$.

Proof.
Let $(F_{ij})_{1\le i,j\le n} : (\Omega, \mathcal{F}, P) \to \{0,1\}^{n\times n}$ be a random double stochastic matrix. Let $\omega \in \Omega$, $F_{ij}(\omega) = f_{ij}$, $1 \le i, j \le n$. Define $\Pi : (\Omega, \mathcal{F}, P) \to S_n$ random permutation such that $\Pi(\omega) = \pi \Leftrightarrow \omega \in \cap_{i,j=1}^{n} F^{-1}_{ij}(f_{ij})$, $\pi(j) := \{i : T_{ij}(\omega) = 1\}$, $1 \le i \le n$. Conversely, define $(T_{ij})_{1\le i,j\le n}$ a random matrix such that $T_{ij}(\omega) = t_{ij} \Leftrightarrow \omega \in \Pi^{-1}(\pi)$, $t_{ij} := 1_{\pi(j)=i}$, $1 \le i, j \le n$. One can easily see that $(T_{ij})_{1\le i,j\le n}$ is indeed a random double stochastic matrix with $0-1$ entries and $\mathbb{T} = \mathbb{F}$ P-a.s. □

Return now to the matrix $\mathbb{Y} = (Y_{ij})_{1\le i,j\le n}$. Let $1 \le r \le n$. Let $\Pi^k : (\Omega, \mathcal{F}, P) \to S_r$ be the random permutations which generate the order statistics for $Y_{k1}, \dots, Y_{kr}$, $1 \le k \le r$, i.e.

$$Y_{k\Pi^k(1)} < \dots < Y_{k\Pi^k(r)}, \quad 1 \le k \le r.$$

By the continuity of the distribution of Y, the random permutations are well defined. Therefore, we can define the random double stochastic matrices $\mathbb{F}^k := (F^k_{ij})_{1\le i,j\le n}$, $1 \le k \le r$,

$$F^k_{ij} := \begin{cases} 1_{\{\Pi^k(j)=i\}} & 1 \le i,j \le r \\ 0 & i,j \notin \{1,...,r\}. \end{cases}$$

Henceforth, if we let r vary, then we denote the matrix above by $(\mathbb{F}^{(r)})^k := (F^{(r)}_{ij})^k_{1\le i,j\le n}$. This is equivalent to $(F^{(r)}_{ij})^k = 1_{\{Y_{kj}=Y_{k(i:r)}\}}$ for $1 \le i,j \le r$ and 0 otherwise.

Definition 41 (Ranked Set Sampling Random Variables) Let $(X_{ij}, Y_{ij})_{1\le i,j\le n}$ be a random matrix of i.i.d. random vectors, each with distribution P^{XY}_θ, $\theta \in \Theta_0$ and let the random double stochastic matrices $(\mathbb{F}^{(r)})^k \in \mathbb{R}^{n\times n}$, $1 \le k \le r$ be $(F^{(r)}_{ij})^k_{1\le i,j\le r} := (1_{\{Y_{kj}=Y_{k(i:r)}\}})_{1\le i,j\le r}$ and 0 otherwise. Then the RSS random variables of size r are sampling random variables $X^{(r)}_{[i]} := (\mathbb{T}^{(r)}_{i*})'\mathbb{X}_{i*}$, $1 \le i \le n$, where the sampling matrix of size r, $\mathbb{T}^{(r)}$ is given by

$$\mathbb{T}^{(r)}_{i*} := \begin{cases} (\mathbb{F}^{(r)}_{i*})^i & 1 \le i \le r \\ (\delta^i)_{1\le j\le n} & r+1 \le i \le n \end{cases}$$

and $\delta^i_{1\le j\le n}$ are vectors satisfying $\delta^i_j = 1 \Leftrightarrow i = j$ and 0 otherwise.

Remark 42 In a notationally more convenient form, the RSS random variables of size r have the form

$$X^{(r)}_{[i]} := \begin{cases} \sum_{j=1}^r 1_{\{Y_{ij}=Y_{i(i:r)}\}} X_{ij} & 1 \le i \le r \\ X_{ii} & r+1 \le i \le n. \end{cases}$$

Proposition 43 Let $\theta \in \Theta_0$. If $1 \le i \le r$, then RSS random variables of size r are distributed as $X^{(r)}_{[i]} \sim P^{X^{(r)}_{[i]}}_\theta$, where

$$P^{X^{(r)}_{[i]}}_\theta(A) = \int_{A\times\mathbb{R}} \frac{1}{B(i,n-i+1)} P^Y_\theta(y)^{i-1}(1-P^Y_\theta(y))^{n-i} P^{X|Y=y}_\theta(dx) P^Y_\theta(dy) \tag{12}$$

for every $A \in \mathcal{B}(\mathbb{R})$. If $r+1 \le i \le n$, then $X^{(r)}_{[i]} \sim P^X_\theta$, where P^X_θ is the marginal probability distribution $P^X_\theta(A) = \int_{A\times\mathbb{R}} P^{XY}_\theta(dx,dy)$, $A \in \mathcal{B}(\mathbb{R})$, $\theta \in \Theta_0$. Moreover, the RSS random variables of size r are independent.

Proof.

Let $\theta \in \Theta_0$ and $A \in \mathcal{B}(\mathbb{R})$. Then, analogously to Lemma 38 we have $P^{X^{(r)}_{[i]}}_\theta(A) =$

$$\begin{aligned} &= \sum_{j=1}^r P(X_{ij} \in A, T^{(r)}_{ij} = 1) \\ &= \sum_{j=1}^r P(X_{ij} \in A, Y_{ij} = Y_{i(i:r)}) \\ &= r\,P(X_{11} \in A, Y_{11} = Y_{1(i:r)}) \\ &= r\,P(X_{11} \in A, \exists!\,(i-1)\text{ of }\{Y_{11},\dots,Y_{1r}\} \setminus \{Y_{1(i:r)}\}\text{ which are } < Y_{1(i:r)}) \end{aligned}$$

$$
\begin{aligned}
&= r\int\int 1_A(x)P(\exists!\,(i-1)\,\mathrm{of}\,\{Y_{11},\ldots,Y_{1r}\}\setminus\{Y_{1(i:r)}\} \\
&\quad \text{which are} < Y_{1(i:r)} \mid Y_{1(i:r)} = y)P_\theta^{XY}(dx,dy) \\
&= r\int\int 1_A(x)\binom{n-1}{i-1}P_\theta^Y(y)^{i-1}(1-P_\theta^Y(y))^{n-i}P_\theta^{X|Y=y}(dx)P_\theta^Y(dy) \\
&= \int\int 1_A(x)\frac{1}{B(i,n-i+1)}P_\theta^Y(y)^{i-1}(1-P_\theta^Y(y))^{n-i}P_\theta^{X|Y=y}(dx)P_\theta^Y(dy)
\end{aligned}
$$

That $X_{[i]}^{(r)} \sim P_\theta^X$, for $r+1 \leq i \leq n$ is obvious. The independence follows by Lemma 38. □

Remark 44 In the case $\{P_\theta^Y : \theta \in \Theta_0\} \ll \lambda$, the distribution of the RSS random variables of size r is determined by

$$
P_\theta^{X_{[i]}^{(r)}}(A) = \int\int 1_A(x)P_\theta^{X|Y=y}(dx)P_\theta^{Y_{(i:r)}}(dy), \quad 1 \leq i \leq r, A \in \mathcal{B}(\mathbb{R}). \tag{13}
$$

Remark 45 Notice also that a single RSS random variable $X_{[i]}^{(r)}$ has the analog probability distribution of a random variable called in literature *concomitant of order statistics* or *induced order statistic* with the difference that a sample of induced order statistics is only conditionally independent. For example, in Bhatacharya [4], the induced order statistics are defined as follows: Let $(Z_1, Y_1), \ldots, (Z_r, Y_r)$ be i.i.d. random vectors, $(Z_i, Y_i) \sim P^{ZY}$, $1 \leq i \leq r$. The i-th induced order statistic $Z_{[i:r]}$ is defined to be $Z_{[i:r]} = Z_j$ if $Y_{(i:r)} = Y_j$, for $1 \leq i \leq r$. In particular, if $P^Y \ll \lambda$, then $Z_{[i:r]}$ has an analog distribution as in equation (13). Despite the fact that each of the random variables in a set of RSS random variables is in fact an induced order statistic, the RSS random variables are due to the sampling separately in each row, independent. In comparison to this, the induced order statistics are derived from a single sample of observations and are conditionally independent given $Y_{(i:r)} = y_i$, $1 \leq i \leq r$: Denote by $P^{Z|Y}$ the conditional distribution of Z given Y. Then it holds

$$
P(Z_{[i:r]} \leq z_i,\, 1 \leq i \leq r \mid Y_1 = y_1, \ldots, Y_r = y_r) = \otimes_{i=1}^r P^{X|Y=y_{(i:r)}}(z_i), \quad z_i, y_i \in \mathbb{R}, 1 \leq i \leq r.
$$

For a proof see Bhatacharya [4], Lemma 3.1. For more details about concomitants of order statistics see also David [11].

Denote in what follows the parameter space $\tilde{\Theta} \subset \Theta_0$ given by

$$
\tilde{\Theta} := \{P_\theta^{XY} \in \Theta_0 : P_\theta^{X|Y=y} = \delta_{\{y\}},\, P_\theta^Y - a.s.\}, \tag{14}
$$

where $\delta_{\{y\}}$ is the Dirac measure. In this case we remark that the distribution of X is the same as the distribution of Y, i.e.

$$
P_\theta^X(A) = \int_{\mathbb{R}^2} 1_A(x)\delta_{\{y\}}(dx)P_\theta^Y(dy) = P_\theta^Y(A), \quad A \in \mathcal{B}(\mathbb{R}), \theta \in \tilde{\Theta}.
$$

Analogously, remark that for this parameter space, the distribution of the i-th RSS random variable is equal to the distribution of the i-th order statistic since

$$\begin{aligned}
P_\theta^{X_{[i]}^{(r)}}(A) &= \int\int 1_A(x)\frac{1}{B(i,n-i+1)}P_\theta^Y(y)^{i-1}(1-P_\theta^Y(y))^{n-i}P_\theta^{X|Y=y}(dx)P_\theta^Y(dy) \\
&= \int\int 1_A(x)\frac{1}{B(i,n-i+1)}P_\theta^X(y)^{i-1}(1-P_\theta^X(y))^{n-i}\delta_{\{y\}}(dx)P_\theta^X(dy) \\
&= \int 1_A(y)\frac{1}{B(i,n-i+1)}P_\theta^X(y)^{i-1}(1-P_\theta^X(y))^{n-i}P_\theta^X(dy) \\
&= \int_A P_\theta^{X_{(i:r)}}(dx)
\end{aligned}$$

for each $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \tilde{\Theta}$.

Definition 46 (Perfect RSS Random Variables) Perfect RSS random variables of size r are RSS random variables of size r $X_{[i]}^{(r)}$, $1 \le i \le n$, with distribution $X_{[i]}^{(r)} \sim P_\theta^{X_{[i]}^{(r)}}$, $\theta \in \tilde{\Theta}$, $1 \le i \le n$.

Remark 47 Notice that we can use the more convenient notational form for the perfect RSS random variables:

$$X_{[i]}^{(r)} := \begin{cases} X_{i(i:r)} & 1 \le i \le r \\ X_{ii} & r+1 \le i \le n. \end{cases}$$

Here $X_{i(i:r)}$ is the i-th order statistic in a sample of size r.

3.3 Ranked Set Sampling Experiments without Repetition

Consider a sample of RSS random variables of size r,

$$X_{[i]}^{(r)} \sim \begin{cases} P_\theta^{X_{[i]}^{(r)}} & 1 \le i \le r \\ P_\theta^X & r+1 \le i \le n \end{cases}$$

for $\theta \in \Theta_0$ and independent.

The probability distribution $P_\theta^{X_{[i]}^{(r)}}$ is determined in equation (12). By independence, the distribution of the vector of RSS random variables of size r, generating a statistical experiment is

$$(X_{[1]}^{(r)}, \ldots, X_{[n]}^{(r)}) \sim \otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X, \quad \theta \in \Theta_0.$$

Now we are able to define the RSS experiments without repetition, which includes the original RSS technique.

Definition 48 (RSS Experiments without Repetition) Let the parameter space $\Theta_0 := \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{ continuous}\}$ and consider parameter spaces $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$.

Then the RSS experiments without repetition for the parameter space Θ_k are a family of statistical experiments generated by RSS random variables

$$G_r^n := (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \Theta_k\}) \in \mathcal{E}(\Theta_k), \quad 1 \le r \le n.$$

Definition 49 (Perfect RSS Experiments) Let the parametrization $\tilde{\Theta}_k \subset \Theta_k$ be $\tilde{\Theta}_k := \{P^{XY} \in \Theta_k : P_\theta^{X|Y=y} = \delta_{\{y\}}, P_\theta^Y - a.s.\}$, $k \in \mathbb{N}$. The perfect RSS experiments without repetition are given by

$$G_r^{n,\text{perfect}} := (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \tilde{\Theta}_k\}), \quad 1 \le r \le n.$$

Remark 50 1) By the properties of the parameter space $\tilde{\Theta}_k$, the perfect RSS experiments are determined by the perfect RSS random variables

$$G_r^{n,\text{perfect}} := (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{(i:r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \tilde{\Theta}_k\}), \quad 1 \le r \le n.$$

2) Notice that the original RSS technique, called in literature the RSS with one cycle, is determined by the statistical experiment G_n^n.
3) As a particular case, the RSS experiment without repetition $G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\})$, $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$, will be called in the sequel the *Simple Random Sampling* (SRS) experiment of size n.
4) Even if the parameter space Θ_k in the definition of the SRS experiment (or the parameter space $\tilde{\Theta}_k$ in the case of perfect RSS experiments) is a subset of the set of all bivariate probability distributions on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R})^2)$, we will not make any further notations and understand the probability distribution P_θ^X (P_θ^Y respectively) as the marginal distribution with respect to X (marginal distribution with respect to Y respectively) of P_θ^{XY}, for $\theta \in \Theta_k$, $k \in \mathbb{N}$.

3.4 Ranked Set Sampling Experiments with Repetition

If $\mathcal{E}_i = (E_i, \mathcal{B}_i, \{P_{\theta,i} : \theta \in \Theta\}) \in \mathcal{E}(\Theta)$, $1 \le i \le n$ are statistical experiments for the parameter space Θ then we denote the product statistical experiment by

$$\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_n := (\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{B}_i, \{\otimes_{i=1}^n P_{\theta,i} : \theta \in \Theta\}).$$

For the definition of a RSS experiment with repetition, we consider a product experiment of RSS experiments without repetition of the full size, where the ranking is done in every row of the original starting matrix and purely random observations are not included any longer. Here, we denote by n the number of repetitions of the experiment, in comparison to the case of RSS without repetition, where n was the notation for largest size of the experiment. This has a motivation for asymptotic considerations of the experiments.

Definition 51 (Unbalanced RSS Experiments with Repetition) Let the parameter space be $\Theta_0 := \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{ continuous}\}$ and consider parameter spaces $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$. Then the unbalanced RSS experiment with n repetitions for the parameter space Θ_k is the product experiment of RSS experiments without repetition

$$\otimes_{i=1}^n G_{t_i}^{t_i} := (\mathbb{R}^{\sum_{i=1}^n t_i}, \mathcal{B}(\mathbb{R})^{\sum_{i=1}^n t_i}, \{\otimes_{i=1}^n \otimes_{j=1}^{t_i} P_\theta^{X_{[j]}^{(t_i)}} \theta \in \Theta_k\}),$$

$t_i \in \mathbb{N}$, $1 \leq i \leq n$.

If $t_1 = \ldots = t_n = p$, $p \in \mathbb{N}$, then the experiment will be called the *balanced RSS experiment with n repetitions*.

Definition 52 (Balanced RSS Experiments with Repetition) Let $\Theta_0 := \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{ continuous}\}$ and consider parameter spaces $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$. Then the balanced RSS experiment with n repetitions for the parameter space Θ_k is the product experiment of RSS experiments without repetition

$$\otimes_{i=1}^n G_p^p := (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^n \otimes_{j=1}^{p} P_\theta^{X_{[j]}^{(p)}} \theta \in \Theta_k\}),$$

$t_i \in \mathbb{N}$, $1 \leq i \leq n$.

Remark 53 Analogously as in the case of RSS experiments without repetition, when restricting to the parameter space $\tilde{\Theta}_k = \{P^{XY} \in \Theta_k : P_\theta^{X|Y=y} = \delta_{\{y\}} P_\theta^Y - a.s.\}$, we obtain the perfect RSS experiment with n repetitions and balanced

$$\otimes_{i=1}^n G_p^{p,perfect} := (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^n \otimes_{j=1}^{p} P_\theta^{X_{(j:p)}} \theta \in \tilde{\Theta}_k\}),$$

$t_i \in \mathbb{N}$, $1 \leq i \leq n$.

4 Comparisons in the family of RSS Experiments without Repetition

In the sequel, the definition of the RSS random variables by means of random stochastic matrices will no longer be used. We will resume our attention only to the RSS experiments as probability spaces which we compare with respect to the informational order in the space of all experiments for a fixed parameter space. This chapter is divided as follows: In the first section of the chapter we recall some facts about sufficiency and completeness and make also some new remarks for this subject. In the second section we derive the non-existence of an informational order between relevant statistical experiments in the RSS problematic. The last section contains examples to motivate the treatment of the comparisons.

4.1 Remarks on Order Statistics, Sufficiency and Completeness

Definition 54 A probability distribution P on $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r)$ is called exchangeable if P is invariant with respect to the group of permutations $\mathcal{S}_r$ on $\{1, \dots, r\}$, i.e. invariant under all maps $\tau_\pi : \mathbb{R}^r \to \mathbb{R}^r$ with

$$\tau_\pi(x_1, \dots, x_r) = (x_{\pi(1)}, \dots, x_{\pi(r)}), \quad \forall \pi \in \mathcal{S}_r. \tag{15}$$

Denote by $\mathcal{B}_0^r := \{B \in \mathcal{B}(\mathbb{R})^r : \tau_\pi B = B \, \forall \pi \in \mathcal{S}_r\} \subset \mathcal{B}(\mathbb{R})^r$ the σ-subfield of τ_π invariant sets.

Proposition 55 If P is an exchangeable probability distribution on $\mathcal{B}(\mathbb{R})^r$ then for all $B \in \mathcal{B}(\mathbb{R})^r$

$$E(1_B \mid \mathcal{B}_0^r) = \frac{1}{r} \sum_{\pi \in \mathcal{S}_r} 1_{\tau_\pi B} \quad P - a.s.$$

i.e. $\mathcal{B}_0^r$ is a sufficient σ-subfield for all exchangeable probability distributions.

Proof.
See Witting [29]. □

Proposition 56 The order statistic $O_r : \mathbb{R}^r \to \mathbb{R}^r$ is sufficient for all exchangeable distributions, i.e.

$$O_r^{-1}\mathcal{B}(\mathbb{R}^r) = \mathcal{B}_0^r.$$

Proof.
We show $O_r^{-1}\mathcal{B}(\mathbb{R})^r = \mathcal{B}_0^r$.
$\subset$: $O_r \circ \tau_\pi = O_r \Rightarrow O_r$ is $\mathcal{B}_0^r$ measurable $\Rightarrow O_r^{-1}\mathcal{B}(\mathbb{R})^r \subset \mathcal{B}_0^r$.
$\supset$: Let $B \in \mathcal{B}_0^r$. If $x \in B$ then we have $O_r x = \tau_{\pi_0}(x) \in B$ for a $\pi_0 \in \mathcal{S}_r$, this implies $x \in O_r^{-1}B \in O_r^{-1}\mathcal{B}(\mathbb{R})^r$. Therefore, $\mathcal{B}_0^r \supset O_r^{-1}\mathcal{B}(\mathbb{R})^r$. □

Corollary 57 Let $\Theta \neq \emptyset$ be an arbitrary parameter set and $\{\otimes_{i=1}^{r} P_\theta^Y : \theta \in \Theta\}$ an arbitrary family of probability distributions on $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r)$. Let O_r be the order statistic on $\mathbb{R}^r$. Then the following affirmations hold:
(1) The experiment $(\mathbb{R}^r, \mathcal{B}_0^r, \{\otimes_{i=1}^{r} P_\theta^Y \mid_{\mathcal{B}_0^r} : \theta \in \Theta\})$ is exhaustive for the experiment $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r, \{\otimes_{i=1}^{r} P_\theta^Y : \theta \in \Theta\})$.
(2) The experiments $(\mathbb{R}^r, \mathcal{B}_0^r, \{(\otimes_{i=1}^{r} P_\theta^Y) \mid_{\mathcal{B}_0^r} : \theta \in \Theta\})$ and $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r, \{P_\theta^{O_r Y^r} : \theta \in \Theta\})$ are mutually exhaustive.
(3) The experiment $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r, \{P_\theta^{O_r Y^r} : \theta \in \Theta\})$ is exhaustive for the experiment $(\mathbb{R}^r, \mathcal{B}(\mathbb{R})^r, \{\otimes_{i=1}^{r} P_\theta^Y : \theta \in \Theta\})$.

Proof.
(1) Denote by $K_1^r : \mathbb{R}^r \times \mathcal{B}(\mathbb{R})^r \to [0,1]$ the Markov kernel

$$K_1^r(\cdot, B) := E_\theta(1_B \mid \mathcal{B}_0^r), \quad \otimes_{i=1}^{r} P_\theta^Y - as. \tag{16}$$

Then by the sufficiency of the σ-algebra $\mathcal{B}_0^r$, the right hand side does not depend a.s. on θ and by the properties of the conditional probability it follows

$$\otimes_{i=1}^{r} P_\theta^Y(B) = \int K_1^r(y_1, \dots, y_r, B)(\otimes_{i=1}^{r} P_\theta^Y) \mid_{\mathcal{B}_0^r} (dy_1, \dots, dy_r)$$

for every $B \in \mathcal{B}(\mathbb{R})^r$ and $\theta \in \Theta$.
(2) Denote by $O_r B := \{O_r(x_1, \dots, x_r) : (x_1, \dots, x_r) \in B\}$ and by $K_2^r : \mathbb{R}^r \times \mathcal{B}_0^r \to [0,1]$ the Markov kernel

$$K_2^r(\cdot, B) := 1_{O_r B}, \quad B \in \mathcal{B}_0^r. \tag{17}$$

Then by Lemma 27 it follows that

$$(\otimes_{i=1}^{r} P_\theta^Y) \mid_{\mathcal{B}_0^r} (B) = \int K_2^r(x_1, \dots, x_r, B) P_\theta^{O_r Y^r}(dx_1, \dots, dx_r)$$

for every $B \in \mathcal{B}_0^r$ and $\theta \in \Theta$.
Conversely, let $K_3^r : \mathbb{R}^n \times \mathcal{B}(\mathbb{R})^r \to [0,1]$ be the Markov kernel

$$K_3^r(\cdot, B) := (1_B \circ O_r).$$

Then also by Lemma 27 it follows that

$$P_\theta^{O_r Y^r}(B) = \int K_3^r(x_1, \dots, x_r, B)(\otimes_{i=1}^{r} P_\theta^Y) \mid_{\mathcal{B}_0^r} (dx_1, \dots, dx_r)$$

for every $B \in \mathcal{B}(\mathbb{R})^r$ and $\theta \in \Theta$.
(3) Define the kernel $K^r : \mathbb{R}^r \times \mathcal{B}(\mathbb{R})^r \to [0,1]$,

$$K^r(\cdot, B) := \int_{\mathbb{R}^r} K_1^r(y_1, \dots, y_r, B) K_2^r(\cdot, dy_1, \dots, dy_r) \tag{18}$$

$$= \int_{\mathbb{R}^r} \frac{1}{r!} \sum_{\pi \in \mathcal{S}_r} 1_{\tau_\pi B}(y_1, \dots, y_r) K_2^r(\cdot, dy_1, \dots, dy_r). \tag{19}$$

Then it follows that

$$\otimes_{i=1}^{r} P_\theta^{Y}(B) = \int_{\mathbb{R}^r} K^r(z_1,\ldots,z_r,B) P_\theta^{O_r Y^r}(dz_1,\ldots,dz_r) \tag{20}$$

for every $B \in \mathcal{B}(\mathbb{R})^r$ and $\theta \in \Theta$. □

For more details see also Proposition 24 and Lemma 27 in the Prerequisites chapter.
Recall the family of RSS experiments without repetition for the parameter space Θ_k, $k \in \mathbb{N}$:

$$G_r^n := (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^{n} P_\theta^{X} : \theta \in \Theta_k\}) \in \mathcal{E}(\Theta_k), \quad 1 \le r \le n,$$

where $\Theta_k \subseteq \Theta_0 := \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{ continuous}\}$, $k \in \mathbb{N}$. In what follows we define parameter spaces Θ_k to take into consideration in the rest of the assertions.

We establish for which families of probabilities distributions, the order statistic is complete. The functions on $\mathbb{R}^r$ depending on $(x_1,\ldots,x_r)$ through the order statistic are the functions of $(x_1,\ldots,x_r)$ which are invariant under all permutations of $(x_1,\ldots,x_r)$. Therefore, completeness of a family $\{P_\theta^{O_r Y^r} : \theta \in \Theta\}$ is the same as symmetrically completeness of order r of $\{P_\theta^{Y} : \theta \in \Theta\}$. The principle to be followed here is stipulated in Proposition 32.

Proposition 58

1. Let $P_0 \in \mathcal{M}(\mathcal{B}(\mathbb{R}))$ and $1 \le q \le \infty$. Then the family of probability distributions

   $$\Delta_1(P_0) := \{P^Y \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : P^Y << P_0,\ dP^Y/dP_0 \text{ bounded}, P^Y \text{with compact support}\}$$

   is q-complete and convex.

2. Let $h_1,\ldots,h_l$ be $\mathcal{B}(\mathbb{R})^m$-measurable non-negative numerical functions. Then the families of probability distributions

   $$\Delta_2 := \{P^Y \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : \int h_i(x_1,\ldots,x_m) \prod_{i=1}^{m} P^Y(dx_i) < \infty,\ i = 1,\ldots,l\},$$

   $$\Delta_3 := \{P^Y \in \Delta_2 : P^Y \text{continuous}\}, \quad \Delta_4 := \{P^Y \in \Delta_2 : P^Y \ll \lambda\},$$

   are q-symmetrically complete of each order n.

Proof.
See Propositions 7.79 and 7.80 in Witting [29].

□

Corollary 59 For the families of probability distributions $\{\otimes_{i=1}^{r} P_\theta^{Y} : \theta \in \Delta_j\}$, $1 \le j \le 4$, the order statistic O_r is complete.

Denote by $\mathcal{M}(\mathcal{B}(\mathbb{R})^2)$ the space of all probability distributions on $\mathcal{B}(\mathbb{R})^2$ and let $\mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) \subseteq \mathcal{M}(\mathcal{B}(\mathbb{R})^2)$. We define the parameter spaces we will consider in the sequel:

$$\begin{aligned}
\Theta_0 &:= \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P^Y \text{continuous}\} &(21)\\
\Theta_1 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^Y \in \Delta_1\} &(22)\\
\Theta_2 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^Y \in \Delta_2\} &(23)\\
\Theta_3 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^Y \in \Delta_3\} &(24)\\
\Theta_4 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^Y \in \Delta_4\} &(25)\\
\Theta_5 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^{XY} \ll \lambda^2, P^Y \in \Delta_2\} &(26)\\
\Theta_6 &:= \{P^{XY} \in \mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) : P^{XY} \ll \lambda^2, P^X \in \Delta_2, P^Y \in \Delta_2\} &(27)
\end{aligned}$$

Proposition 60 If $\mathcal{P} := \{P_\theta^{O_r Y^r} : \theta \in \Theta\}$ is a complete family of probability distributions then the families $\mathcal{P}_i := \{P_\theta^{Y_{(i:r)}} : \theta \in \Theta\}$, $1 \leq i \leq r$ are also complete.

Proof.
Let $h \in \bigcap_{\theta\in\Theta} L_1(P_\theta^{Y_{(i:r)}})$ such that

$$\int h(y) P_\theta^{Y_{(i:r)}}(dy) = 0, \quad \theta \in \Theta.$$

But this is equivalent to $\int h(y_i) P_\theta^{O_r Y^r}(dy_1, \ldots, dy_r) = 0$, $\theta \in \Theta$. Since O_r is complete for this family of probability distributions, it follows that $g = 0$, $P_\theta^{O_r Y^r}$-a.s. where $g : \mathbb{R}^r \to \mathbb{R}$ is defined by $(y_1, \ldots, y_r) \mapsto h(y_i)$ for all $(y_1, \ldots, y_r) \in \mathbb{R}^r$. This implies $h = 0$ $P_\theta^{Y_{(i:r)}}$-a.s. which proves the assertion. □

Remark 61 The order statistic O_r is minimal sufficient for the family of probability distributions $\{\otimes_{i=1}^r P_\theta^Y : \theta \in \Delta_j\}$, $1 \leq j \leq 4$. This is a consequence of Proposition 30. Therefore, the kernel K^r defined in equation (18) is a version of the conditional probability distribution $P^{O_r Y^r | O_r(x_1,\ldots,x_r) = y_1,\ldots,y_r}$. This follows from Proposition 26.

Proposition 62 Let $\{\otimes_{i=1}^r P_\theta^Y : \theta \in \Theta\}$ be a symmetrically complete family of probability distributions. Then, for all $B \in \mathcal{B}(\mathbb{R})^r$, $B := \underbrace{\mathbb{R} \times \ldots \times \mathbb{R} \times}_{\text{i-1 times}} A \underbrace{\times \mathbb{R} \times \ldots \times \mathbb{R}}_{\text{r-i-1 times}}$, $A \in \mathcal{B}(\mathbb{R})$, the following equation holds:

$$K^r(\cdot, B) = \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)} \mathbb{R}\times\ldots\times\mathbb{R}\times A\times\mathbb{R}\times\ldots\times\mathbb{R}}, \quad P_\theta^{O_r Y^r} - a.s., \theta \in \Theta, \tag{28}$$

where we have denoted by $\tau_{(ij)} : \mathbb{R}^r \to \mathbb{R}^r$ the map τ_π restricted to all transpositions $\pi = (i,j) \in \mathcal{S}_r$, $1 \leq i, j \leq r$.

Proof.
Let $B \in \mathcal{B}(\mathbb{R})^r$ such that $B = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R} \times}_{\text{i-1 times}} A \underbrace{\times \mathbb{R} \times \ldots \times \mathbb{R}}_{\text{r-i-1 times}}$, $A \in \mathcal{B}(\mathbb{R})$. Then by equation (18) it follows that $P_\theta^Y(A) = \otimes_{i=1}^r P_\theta^Y(B) =$

$$
\begin{aligned}
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \frac{1}{r!} \sum_{\pi \in \mathcal{S}_n} 1_{\tau_\pi B}(y_1, \ldots, y_r) K_2^r(x_1, \ldots, x_r, dy_1, \ldots, dy_r) P_\theta^{O_r Y^r}(dx_1, \ldots, dx_r) \\
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \frac{1}{r!} \sum_{\pi \in \mathcal{S}_n} 1_{\tau_\pi \mathbb{R} \times \ldots \times A \times \ldots \times \mathbb{R}}(y_1, \ldots, y_r) K_2^r(x_1, \ldots, x_r, dy_1, \ldots, dy_r) P_\theta^{O_r Y^r}(dx_1, \ldots, dx_r) \\
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)}(\mathbb{R} \times \ldots \mathbb{R} \times A \times \mathbb{R} \times \ldots \times \mathbb{R})}(y_1, \ldots, y_r) K_2^r(x_1, \ldots, x_r, dy_1, \ldots, dy_r) \\
&\quad P_\theta^{O_r Y^r}(dx_1, \ldots, dx_r).
\end{aligned}
$$

Additionally, by the fundamental equation (11) we also have that:

$$
\begin{aligned}
P_\theta^Y(A) &= \frac{1}{r} \sum_{i=1}^r P_\theta^{Y_{(i:r)}}(A) \\
&= \frac{1}{r} \sum_{i=1}^r \int_{\mathbb{R}} 1_A(x) P_\theta^{Y_{(i:r)}}(dx) \\
&= \int_{\mathbb{R}^r} \frac{1}{r} \sum_{i=1}^r 1_A(x_i) \prod_{i=1}^r P_\theta^{Y_{(i:r)}}(dx_i) \\
&= \int_{\mathbb{R}^r} \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)}(\mathbb{R} \times \ldots \mathbb{R} \times A \times \mathbb{R} \times \ldots \times \mathbb{R})}(x_1, \ldots, x_r) \prod_{i=1}^r P_\theta^{Y_{(i:r)}}(dx_i) \\
&= \int_{\mathbb{R}^r} \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)}(\mathbb{R} \times \ldots \mathbb{R} \times A \times \mathbb{R} \times \ldots \times \mathbb{R})}(x_1, \ldots, x_r) P_\theta^{O_r Y^r}(dx_1, \ldots, dx_r).
\end{aligned}
$$

Therefore,

$$
\int_{\mathbb{R}^r} \left(K^r(x_1, \ldots, x_r, B) - \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)}(B)}(x_1, \ldots, x_r) \right) P_\theta^{O_r Y^r}(dx_1, \ldots, dx_r) = 0,
$$

for every $\theta \in \Theta$ and $B \in \mathcal{B}(\mathbb{R})^r$ with the specified form. The completeness of the order statistic implies that

$$
K^r(\cdot, B) = \frac{1}{r} \sum_{j=1}^r 1_{\tau_{(ij)} B},\ P_\theta^{O_r Y^r} - a.s.
$$

for every $\theta \in \Theta$ and $B \in \mathcal{B}(\mathbb{R})^r$ with the specified form. The assertion is proved.

□

4.2 The Informational Order

The main purposes of this section is to prove, that the family of RSS experiments without repetition,

$$G_r^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \Theta_k\}), \quad 1 \le r \le n, k \in \{5, 6\}$$

is not more informative, i.e. $\exists r \in \{1, \ldots, n-1\}$ such that G_{r+1}^n is not more informative than G_r^n, and perhaps the most relevant result, that G_n^n is not more informative than G_1^n, in other words, the original RSS experiment is not more informative than the SRS experiment. In order to prove the non-existence of the informational order mentioned above, we restrict first our attention to statistical experiments on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ generated by a single RSS random variable. These experiments can also be viewed as experiments generated by concomitants of order statistics. Although the non-existence of the informational order is proved for the parameter spaces Θ_5 and Θ_6 (here we need dominated experiments) defined in equations (26) and (27), important results regarding the exhaustivity of RSS experiments hold for the parameter spaces Θ_k, $1 \le k \le 6$.
The motivation of the next theorem is the following: When observing a RSS random variable of size $r+1$, one could believe that it contains more information than a RSS random variable of size r. This because the first is arising from the ordering of the larger sample $Y_1, \ldots, Y_{r+1}$. We prove that, contrary to the intuition, the affirmation does not hold.

Theorem 63 Let $1 \le k \le 6$. Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$ there exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_\cdot^{X|Y}$. Then for every $1 \le i \le r$ and $1 \le j \le r+1$ the experiment

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^{X_{[j]}^{(r+1)}} : \theta \in \Theta_k\})$$

is not exhaustive for the experiment

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^{X_{[i]}^{(r)}} : \theta \in \Theta_k\}).$$

Proof.
We will prove the assertion by a contradiction argument. We assume the exhaustivity of the experiments takes place, i.e. we assume there exists a Markov kernel L from $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$P_\theta^{X_{[i]}^{(r)}}(A) = \int_{\mathbb{R}} L(x, A) P_\theta^{X_{[j]}^{(r+1)}}(dx), \quad \theta \in \Theta_k, A \in \mathcal{B}(\mathbb{R}). \tag{29}$$

For the right-hand side of equation (29) we have then

$$\int_{\mathbb{R}} L(x, A) P_\theta^{X_{[j]}^{(r+1)}}(dx) = \int_{\mathbb{R}} \int_{\mathbb{R}} L(x, A) P_\cdot^{X|Y=y}(dx) P_\theta^{Y_{(j:r+1)}}(dy)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}} L(x,A) P_{\cdot}^{X|Y=y_j}(dx) P_\theta^{O_{r+1}Y^{r+1}}(dy_1,\dots,dy_{r+1}) \\
&= \int_{\mathbb{R}^{r+1}} g_A(y_1,\dots,y_{r+1}) P_\theta^{O_{r+1}Y^{r+1}}(dy_1,\dots,dy_{r+1}),
\end{aligned}
$$

where we have denoted by $g_A : \mathbb{R} \to \mathbb{R}$ the measurable function

$$
(y_1,\dots,y_{r+1}) \mapsto \int_{\mathbb{R}} L(x,A) P_{\cdot}^{X|Y=y_j}(dx). \tag{30}
$$

For the left-hand side of equation (29) it follows that:

$$
\begin{aligned}
P_\theta^{X_{[i]}^{(r)}}(A) &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x) P_{\cdot}^{X|Y=t}(dx) P_\theta^{Y_{(i:r)}}(dt) \\
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) P_{\cdot}^{X|Y=t_i}(dx) P_\theta^{O_r Y^r}(dt_1\dots,dt_r) \\
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) \frac{dP_\theta^{O_r Y^r}}{d(P_\theta^Y)^r}(t_1,\dots,t_r) P_{\cdot}^{X|Y=t_i}(dx) \prod_{i=1}^r P_\theta^Y(dt_i) \\
&= \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) r! 1_{t_1<\ldots<t_r} P_{\cdot}^{X|Y=t_i}(dx) \prod_{i=1}^r P_\theta^Y(dt_i) \\
&= \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}} 1_A(x) r! 1_{t_1<\ldots<t_r} P_{\cdot}^{X|Y=t_i}(dx) \prod_{i=1}^{r+1} P_\theta^Y(dt_i) \\
&= \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}} 1_A(x) r! 1_{t_1<\ldots<t_r} P_{\cdot}^{X|Y=t_i}(dx) \\
&\quad K^{r+1}(y_1,\dots,y_{r+1},dt_1,\dots,dt_{r+1}) P_\theta^{O_{r+1}Y^{r+1}}(dy_1,\dots,dy_{r+1}) \\
&= \int_{\mathbb{R}^{r+1}} h_A(y_1,\dots,y_{r+1}) P_\theta^{O_{r+1}Y^{r+1}}(dy_1,\dots,dy_{r+1})
\end{aligned}
$$

where K^{r+1} is the Markov kernel defined in equation (18), and where we have denoted by $h_A : \mathbb{R}^{r+1} \to \mathbb{R}$ the measurable function

$$
(y_1,\dots,y_{r+1}) \mapsto \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}} 1_A(x) r! 1_{\{t_1<\ldots<t_r\}} P_{\cdot}^{X|Y=t_i}(dx) K^{r+1}(y_1,\dots,y_{r+1},dt_1,\dots,dt_{r+1})
$$

for every $A \in \mathcal{B}(\mathbb{R})$. Also, by equation (18) it follows that we can rewrite the function h_A above as

$$
\begin{aligned}
&\int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}^{r+1}} \int_{\mathbb{R}} 1_A(x) r! 1_{\{t_1<\ldots<t_r\}} P_{\cdot}^{X|Y=t_i}(dx) \left(\frac{1}{(r+1)!} \sum_{\pi \in \mathcal{S}_{r+1}} \delta_{\{z_1,\dots,z_{r+1}\}}(\tau_\pi(dt_1,\dots,dt_{r+1}) \right) \\
&K_2^{r+1}(y_1,\dots,y_{r+1},dz_1,\dots,dz_{r+1})
\end{aligned}
$$

where the Markov kernel $K_2^{r+1} : \mathbb{R}^{r+1} \times \mathcal{B}(\mathbb{R})^{r+1} \to [0,1]$ is given by $K_2^{r+1}(\cdot, B) = 1_{O_{r+1}B}$ for every $\mathcal{B} \in \mathcal{B}(\mathbb{R})^{r+1}$.

Therefore, h_A is equal to

$$\begin{aligned}
& \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{\{t_1<\ldots<t_r\}}P_{\cdot}^{X|Y=t_i}(dx)\delta_{\{z_1,\ldots,z_{r+1}\}}\tau_\pi(dt_1,\ldots,dt_{r+1}) \\
& K_2^{r+1}(y_1,\ldots,y_{r+1},dz_1,\ldots,dz_{r+1}) \\
= & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{\{t_1<\ldots<t_r\}}P_{\cdot}^{X|Y=t_i}(dx)\delta_{\tau_\pi(\{z_1,\ldots,z_{r+1}\})}(dt_1,\ldots,dt_{r+1}) \\
& K_2^{r+1}(y_1,\ldots,y_{r+1},dz_1,\ldots,dz_{r+1}) \\
= & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{\{z_{\pi(1)}<\ldots<z_{\pi(r)}\}}P_{\cdot}^{X|Y=z_{\pi(i)}}(dx)\delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1}) \\
= & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{\{z_{\pi(1)}<\ldots<z_{\pi(r)}<z_{\pi(r+1)}\}}P_{\cdot}^{X|Y=z_{\pi(i)}}(dx) \\
& \delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1}) + \\
+ & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{B_\pi}(z_1,\ldots,z_{r+1})P_{\cdot}^{X|Y=z_{\pi(i)}}(dx)\delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1})
\end{aligned}$$

where we have denoted by $B_\pi\in\mathcal{B}(\mathbb{R})^{r+1}$ the set such that $\{z_{\pi(1)}<\ldots<z_{\pi(r)}<z_{\pi(r+1)}\}\cup B_\pi=\{z_{\pi(1)}<\ldots<z_{\pi(r)}\}$ and $\{z_{\pi(1)}<\ldots<z_{\pi(r)}<z_{\pi(r+1)}\}\cap B_\pi=\emptyset$, for every $\pi\in\mathcal{S}_{r+1}$. In the first term of the sum above we can proceed easily with the integration, hence $h_A(y_1,\ldots,y_{r+1})=$

$$\begin{aligned}
= & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}} 1_A(x)P_{\cdot}^{X|Y=y_{\pi(i)}}(dx) + \\
+ & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{B_\pi}(z_1,\ldots,z_{r+1})P_{\cdot}^{X|Y=z_{\pi(i)}}(dx)\delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1}) \\
= & \frac{1}{r+1}\sum_{i=1}^{r+1}\int_{R} 1_A(x)P_{\cdot}^{X|Y=y_i}(dx) + \\
+ & \frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{B_\pi}(z_1,\ldots,z_{r+1})P_{\cdot}^{X|Y=z_{\pi(i)}}(dx)\delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1}).
\end{aligned}$$

Summarizing, we obtain that

$$\int_{\mathbb{R}^{r+1}}(g_A(y_1,\ldots,y_{r+1})-h_A(y_1,\ldots,y_{r+1}))P_\theta^{O_{r+1}Y^{r+1}}(dy_1,\ldots,dy_{r+1})=0,$$

for all $\theta\in\Theta_k$ and $A\in\mathcal{B}$. Remark also that $(g_A-h_A)\in\bigcap_{\theta\in\Theta_k}L_1(P_\theta^{O_{r+1}Y^{r+1}})$, for every $A\in\mathcal{B}(\mathbb{R})$. Moreover, the order statistic is complete for the parameter space Θ_k,

$1 \leq k \leq 6$, it follows that $g_A = h_A$, $P_\theta^{O_{r+1}Y^{r+1}} - as.$ for every fixed but arbitrary $A \in \mathcal{B}(\mathbb{R})$. Since $h_A(y_1, \ldots, y_{r+1}) = \frac{1}{r+1}\sum_{i=1}^{r+1}\int_R 1_A(x)P_\cdot^{X|Y=y_i}(dx) + r_A$, where we have denoted by r_A the function on $\mathbb{R}^{r+1}$, $(y_1, \ldots, y_{r+1}) \mapsto$

$$\frac{1}{r+1}\sum_{\pi\in\mathcal{S}_{r+1}}\int_{\mathbb{R}^{r+1}}\int_{\mathbb{R}} 1_A(x)1_{B_\pi}(z_1,\ldots,z_{r+1})P_\cdot^{X|Y=z_{\pi(i)}}(dx)\delta_{\{y_1,\ldots,y_{r+1}\}}O_{r+1}(dz_1,\ldots,dz_{r+1}),$$

we remark that h_A depends on y_i, $1 \leq i \leq r+1$. This implies that g_A depends on y_i, $1 \leq i \leq r+1$, fact which by the definition of the function g_A, certainly leads to a contradiction. □

Remark 64 We can always find $\mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) \subseteq \mathcal{M}(\mathcal{B}(\mathbb{R})^2)$ such that for Θ_k, $1 \leq k \leq 6$ it exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$. The most natural example would be the restriction to the perfect RSS experiments, i.e. we define $\mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) = \{P^{XY} \in \mathcal{M}(\mathcal{B}(\mathbb{R})^2) : P_\cdot^{X|Y=y} = \delta_{\{y\}}, P^Y - a.s.\}$. We can also relax the requirement of the existence of the conditional distribution which is independent of $\theta \in \Theta_k$ and only assume the existence of a conditional distribution which is independent of the family of the Y marginal distributions. For example consider the family of bivariate distributions

$$\mathcal{M}_0(\mathcal{B}(\mathbb{R})^2) := \{\left(\frac{1}{2\pi\sigma\tau}e^{-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\frac{(x-y)^2}{\tau^2}\right)}\right)\lambda^2 : \sigma,\tau > 0\}.$$

For this family of distributions we have $Y \sim N(0,\sigma^2)$ and the conditional distribution $P^{X|Y=y} = N(y,\tau^2)$ $P^Y - a.s.$, therefore, independent of σ^2. For this family of distributions though, the order statistic is no longer complete, but instead of the completeness of the order statistic, we could use the completeness property in the families of exponential distributions and proceed with the proof in a similar way as above.

Proposition 65 Let $1 \leq k \leq 6$. Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$ there exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_\cdot^{X|Y}$. Then the statistical experiment generated by a RSS random variable

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^{X_{[i]}^{(r)}} : \theta \in \Theta_k\}),$$

$1 \leq i \leq r$, is not exhaustive for the statistical experiment

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
We will prove the assertion by a contradiction argument. We assume the exhaustivity of the experiments, i.e. we assume that there exists a Markov kernel $L : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ such that

$$P_\theta^X(A) = \int_{\mathbb{R}} L(x,A)P_\theta^{X_{[i]}^{(r)}}(dx) \tag{31}$$

for every $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$. Then for the right-hand side of equation (31) we have

$$\int_{\mathbb{R}} L(x,A) P_\theta^{X_{[i]}^{(r)}}(dx) =$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} L(x,A) P_\cdot^{X|Y=y}(dx) P_\theta^{Y_{(i:r)}}(dy)$$

$$\begin{aligned} &= \int_{\mathbb{R}^r} \left(\int_{\mathbb{R}} L(x,A) P_\cdot^{X|Y=y_i}(dx) \right) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r) \\ &= \int_{\mathbb{R}^r} g_A(y_1,\ldots,y_r) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r) \end{aligned}$$

where we have denoted for each $A \in \mathcal{B}(\mathbb{R})$ by $g_A : \mathbb{R}^n \to \mathbb{R}$ the measurable function

$$(y_1,\ldots,y_r) \mapsto \int_{\mathbb{R}} L(x,A) P_\cdot^{X|Y=y_i}(dx).$$

For the left hand-side of equation (31) we have $P_\theta^X(A) =$

$$\begin{aligned} &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x) P_\theta^{XY}(dx,dt) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=t}(dx) P_\theta^Y(dt) \\ &= \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=t_i}(dx) \prod_{i=1}^r P_\theta^Y(dt_i) \\ &= \int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=t_i}(dx) K^r(y_1,\ldots,y_r,dt_1,\ldots,dt_r) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r), \end{aligned}$$

where K^r is the Markov kernel determined in equation (18). Now, by Proposition 62 it follows that the integration with respect to K^r reduces to

$$\begin{aligned} &\int_{\mathbb{R}^r} \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=t_i}(dx) \left(\frac{1}{r!} \sum_{\pi \in \mathcal{S}_r} 1_{\tau_\pi(dt_1,\ldots dt_r)}(y_1,\ldots,y_r) \right) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}^r} \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=t_i}(dx) \left(\frac{1}{r} \sum_{j=1}^r \delta_{\mathrm{pr}_i \tau_{(ij)}(y_1,\ldots,y_r)}(dt_i) \right) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r) \\ &= \int_{\mathbb{R}^r} \left(\frac{1}{r} \sum_{j=1}^r \int_{\mathbb{R}} 1_A(x) P_\cdot^{X|Y=y_j}(dx) \right) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r) \\ &= \int_{\mathbb{R}^r} \left(\tilde{g}_A(y_1,\ldots,y_r) \right) P_\theta^{O_r Y^r}(dy_1,\ldots,dy_r), \end{aligned}$$

where we have denoted for each $A \in \mathcal{B}(\mathbb{R})$ by $\tilde{g}_A : \mathbb{R}^r \to \mathbb{R}$ the measurable function

$$(y_1, \ldots, y_r) \mapsto \frac{1}{r} \sum_{j=1}^{r} \int_{\mathbb{R}} 1_A(x) P_{\cdot}^{X|Y=y_j}.$$

For each $A \in \mathcal{B}(\mathbb{R})$ the functions $g_A, \tilde{g}_A \in \bigcap_{\theta \in \Theta_k} L_1(P_\theta^{O_r Y^r})$. Notice also that the order statistic is for the particular choice of the parameter sets complete, therefore $g_A = \tilde{g}_A$, $P_\theta^{O_r Y^r} - a.s.$, for every $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$. Since the function $\tilde{g}_A$ depends on y_i, $1 \leq i \leq r$ this implies that g_A also depends on y_i, $1 \leq i \leq r$. But this leads to a contradiction since by definition, the function g_A is independent of y_j, $1 \leq j \neq i \leq r$. □

The next proposition is relevant for example in cases where we consider decision functions from the full RSS experiments which are defined for every sample size, i.e. $\rho^n : R^n \times \mathcal{B}(D) \to [0,1]$. One possible question in this case is: does the risk of this decision function decrease while the sample size increases or is it possible that there are situations where a smaller sample size is more informative? The answer to this question can be given also in terms of the exhaustivity of statistical experiments.

Proposition 66 Let $1 \leq k \leq 6$. Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$ there exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_{\cdot}^{X|Y}$. Then the RSS experiment

$$G_{n+1}^{n+1} = (\mathbb{R}^{n+1}, \mathcal{B}(\mathbb{R})^{n+1}, \{\otimes_{i=1}^{n+1} P_\theta^{X_{[i]}^{(n+1)}} : \theta \in \Theta_k\})$$

is exhaustive for the experiment

$$G_n^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^{n} P_\theta^{X_{[i]}^{(n)}} : \theta \in \Theta_k\})$$

if and only if $n = 1$.

$(n = 1) \Rightarrow$ Exhaustivity: Treated in Proposition 84.
Exhaustivity $\Rightarrow (n = 1)$: We assume there exists a Markov kernel $L : \mathbb{R}^{n+1} \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ such that the exhaustivity condition is fulfilled

$$\otimes_{i=1}^{n} P_\theta^{X_{[i]}^{(n)}}(B) = \int_{\mathbb{R}^{n+1}} L(x_1, \ldots, x_{n+1}; B) \prod_{i=1}^{n+1} P_\theta^{X_{[i]}^{(n+1)}}(dx_i) \tag{32}$$

for every $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$. We prove that the exhaustivity condition can take place only when $n = 1$. Making use of the proposition of Fubini, the left-hand side of the equation (32) is then equal to

$$
\begin{aligned}
& \int_{\mathbb{R}^{n+1}} \int_{\mathbb{R}^{n+1}} L(x_1,\dots,x_{n+1};B) \prod_{i=1}^{n+1} P_{\cdot}^{X|Y=y_i}(dx_i) \prod_{i=1}^{Y_{(i:n+1)}} (dy_i) \\
= & \int_{\mathbb{R}^{(n+1)^2}} \int_{\mathbb{R}^{n+1}} L(x_1,\dots,x_{n+1};B) \prod_{i=1}^{n+1} P_{\cdot}^{X|Y=y_i^i}(dx_i) \prod_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}}(dy_1^i,\dots,dy_{n+1}^i) \\
= & \int_{\mathbb{R}^{(n+1)^2}} g_B(y_j^i, 1\le i,j\le n+1) \prod_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}}(dy_1^i,\dots,dy_{n+1}^i)
\end{aligned}
$$

where for each $B \in \mathcal{B}(\mathbb{R})^n$ we have denoted by $g_B : \mathbb{R}^{(n+1)^2} \longrightarrow \mathbb{R}$ the measurable function

$$
(y_j^i : 1 \le i,j \le n+1) \mapsto \int_{\mathbb{R}^{n+1}} L(x_1,\dots,x_{n+1};B) \prod_{i=1}^{n+1} P_{\cdot}^{X|Y=y_i^i}(dx_i).
$$

On the other hand, we know from Proposition 63 that the following equation holds

$$
P_\theta^{X_{[i]}^{(n)}}(A) = \int_{\mathbb{R}^{n+1}} h_A(y_1,\dots,y_{n+1}) P_\theta^{O_{n+1}Y^{n+1}}(dy_1,\dots,dy_{n+1})
$$

for every $1 \le i \le n$, $A \in \mathcal{B}(\mathbb{R})$, $\theta \in \Theta_k$ and where the function $h_A : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ is given by $h_A(y_1,\dots,y_{n+1}) =$

$$
\frac{1}{n+1} \sum_{\pi \in \mathcal{S}_{n+1}} \int_{\mathbb{R}} \int_{\mathbb{R}^{n+1}} 1_A(x) 1_{\{z_{\pi(1)} < \dots < z_{\pi(n)}\}} P_{\cdot}^{X|Y=z_{\pi(i)}}(dx) \delta_{\{y_1,\dots,y_{n+1}\}} O_{n+1}(dz_1,\dots,dz_{n+1})
$$

for every $A \in \mathcal{B}(\mathbb{R})$.
Therefore for every $\theta \in \Theta_k$, $B \in \mathcal{B}(\mathbb{R})^r$, $B = B_1 \times \dots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$ the left-hand side of equation (32) is determined by

$$
\begin{aligned}
\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(B) &= \prod_{i=1}^n \int_{\mathbb{R}} 1_{B_i}(x) P_\theta^{X_{[i]}^{(n)}}(dx) \\
&= \prod_{i=1}^n \int_{\mathbb{R}^{n+1}} h_{B_i}(y_1,\dots,y_{n+1}) P_\theta^{O_{n+1}Y^{n+1}}(dy_1,\dots,dy_{n+1}) \\
&= \int_{\mathbb{R}^{n(n+1)}} \prod_{i=1}^n h_{B_i}(y_1^i,\dots,y_{n+1}^i) \prod_{i=1}^n P_\theta^{O_{n+1}Y^{n+1}}(dy_1^i,\dots,dy_{n+1}^i) \\
&= \int_{\mathbb{R}^{(n+1)^2}} \prod_{i=1}^n h_{B_i}(y_1^i,\dots,y_{n+1}^i) \prod_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}}(dy_1^i,\dots,dy_{n+1}^i) \\
&= \int_{\mathbb{R}^{(n+1)^2}} \tilde{g}_B(y_j^i : 1 \le i,j \le n+1) \prod_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}}(dy_1^i,\dots,dy_{n+1}^i)
\end{aligned}
$$

where for each $B \in \mathcal{B}(\mathbb{R})^n$, $B = B_1 \times \cdots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$ we have defined the measurable function $\tilde{g}_B : \mathbb{R}^{(n+1)^2} \to \mathbb{R}$ by

$$(y_j^i : 1 \leq i, j \leq n+1) \mapsto \prod_{i=1}^{n} h_{B_i}(y_1^i, \ldots, y_{n+1}^i).$$

Notice that for a fixed but arbitrary $B \in \mathcal{B}(\mathbb{R})^n$, $B = B_1 \times \cdots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$, the functions $g_B, \tilde{g}_B \in \bigcap_{\theta \in \Theta_k} L_1(\otimes_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}})$. Now, the order statistic is complete for the chosen parametric families, therefore, by Proposition 31 it follows that g_B and $\tilde{g}_B$ are equal $\otimes_{i=1}^{n+1} P_\theta^{O_{n+1}Y^{n+1}}$-a.s. and for every $\theta \in \Theta_k$ which is possible only in case that $\tilde{g}_B$ is independent of the coordinates $(y_j^i : 1 \leq i \neq j \leq n+1)$. But this is again possible only in the case $n = 1$, i.e. when

$$\tilde{g}_B(y_j^i : 1 \leq i, j \leq 2) = h_B(y_1^1, y_2^1) = \frac{1}{2} \sum_{j=1}^{2} \int 1_B(x) P_\cdot^{X|Y=y_j^1}(dx), \quad B \in \mathcal{B}(\mathbb{R}).$$

□

Proposition 67 Let $1 \leq k \leq 6$. Assume that for the family of probability distributions Θ_k, $1 \leq k \leq 6$ it exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_\cdot^{X|Y}$. Then the RSS experiment without repetition of full size

$$G_n^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} : \theta \in \Theta_k\})$$

is not exhaustive for the SRS experiment of size n

$$G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
Let $1 \leq k \leq 6$. We prove the assertion of the proposition by a contradiction argument. We assume there exists a Markov kernel $L : \mathbb{R}^n \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ such that the exhaustivity condition is fulfilled:

$$\otimes_{i=1}^n P_\theta^X(B) = \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{j=1}^{n} P_\theta^{X_{[j]}^{(n)}}(dx_j) \tag{33}$$

for every $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$. By using the proposition of Fubini, the right-hand side of equation (33) is equal to

$$\begin{aligned} &\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{j=1}^{n} P_\cdot^{X|Y=y_j}(dx_j) P_\theta^{Y_{(j:n)}}(dy_j) \\ = &\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{j=1}^{n} P_\cdot^{X|Y=y_j}(dx_j) \prod_{j=1}^{n} P_\theta^{Y_{(j:n)}}(dy_j) \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{n^2}} \left(\int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^j}(dx_j) \right) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\theta^{Y_{(j:n)}}(dy_j^i) \\
&= \int_{\mathbb{R}^{n^2}} g_B(y_j^i : 1 \le i, j \le n) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\theta^{Y_{(j:n)}}(dy_j^i).
\end{aligned}
$$

The last implication follows after the multiplication of the equation by $\int_{\mathbb{R}} P_\theta^{Y_{(j:n)}}(dy_j^i)$, for $1 \le i \ne j \le n$, i.e. by multiplication with the identity. For each $B \in \mathcal{B}(\mathbb{R})^n$ we have denoted by $g_B : \mathbb{R}^{n^2} \to \mathbb{R}$ the measurable function

$$
(y_j^i : 1 \le i, j \le n) \mapsto \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^j}(dx_j). \tag{34}
$$

By Proposition 84, the statistical experiment $(\mathbb{R}^{n^2}, \mathcal{B}(\mathbb{R})^{n^2}, \{\otimes_{i=1}^n \otimes_{j=1}^n P_\theta^{X_{[j]}^{(n)}} : \theta \in \Theta_k\})$ is exhaustive for the experiment $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\})$, and the following equation holds

$$
\otimes_{i=1}^n P_\theta^X(B) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \prod_{i=1}^{n} K^n(x_1^i, \ldots, x_n^i, B_i) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\theta^{X_{[j]}^{(n)}}(dx_j^i)
$$

for every $B \in \mathcal{B}(\mathbb{R})^n$, $B = B_1 \times \cdots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$. Here $K^n : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ is the Markov kernel

$$
K^n(x_1, \ldots, x_n, A) = \frac{1}{n} \sum_{j=1}^{n} 1_{\tau_{(1j)}(A \times \mathbb{R} \times \ldots \times \mathbb{R})}(x_1, \ldots, x_n), \quad A \in \mathcal{B}(\mathbb{R}).
$$

Let $B \in \mathcal{B}(\mathbb{R})^n$ such that $B = B_1 \times \cdots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$. Then the left-hand side of the equation (33), after the analogous multiplication by 1, is equal to

$$
\begin{aligned}
&= \int_{\mathbb{R}^{n^2}} \left(\int_{\mathbb{R}^{n^2}} \prod_{i=1}^{n} K^n(x_1^i, \ldots, x_n^i, B_i) \prod_{i=1}^{n} \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(dx_j^i) \right) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\theta^{Y_{(j:n)}}(dy_j^i) \\
&= \int_{\mathbb{R}^{n^2}} \tilde{g}_B(y_j^i : 1 \le i, j \le n) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\theta^{Y_{(j:n)}}(dy_j^i)
\end{aligned}
$$

where we have denoted by $\tilde{g}_B : \mathbb{R}^{n^2} \to \mathbb{R}$ the measurable function $\tilde{g}_B(y_j^i : 1 \le i, j \le n) =$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{n^2}} \prod_{i=1}^{n} K^n(x_1^i, \ldots, x_n^i, B_i) \prod_{i=1}^{n} \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(dx_j^i) \\
&= \prod_{i=1}^{n} \int_{R^n} K^n(x_1^i, \ldots, x_n^i, B_i) \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(dx_j^i)
\end{aligned}
$$

$$= \prod_{i=1}^{n} \int_{R^n} \frac{1}{n} \sum_{j=1}^{n} 1_{\tau_{(1j)}(B_i\times\mathbb{R}\times\ldots\times\mathbb{R})}(x_1,\ldots,x_n) \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(dx_j^i)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} \int_{\tau_{(1j)}(B_i\times\mathbb{R}\times\ldots\times\mathbb{R})} \prod_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(dx_j^i) \right)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{n} \sum_{j=1}^{n} P_{\cdot}^{X|Y=y_j^i}(B_i) \right).$$

For every $B \in \mathcal{B}(\mathbb{R})^n$ the functions g_B and $\tilde{g}_B$ belong to $\bigcap_{\theta\in\Theta_k} L_1((\otimes_{j=1}^n P_\theta^{Y_{(j:n)}})^n)$. Therefore, by Lemma 31 it follows that $g_B = \tilde{g}_B$ $(\otimes_{j=1}^n P_\theta^{Y_{(j:n)}})^n$-a.s., $\forall \theta \in \Theta_k$ and for every $B \in \mathcal{B}(\mathbb{R})^n$ such that $B = B_1 \times \cdots \times B_n$, $B_i \in \mathcal{B}(\mathbb{R})$. This implies that the function g_B depends as $\tilde{g}_B$, on all y_j^i, $1 \le i, j \le n$ which by the definition of g_B in equation (34) leads to a contradiction. $\square$

Proposition 68 Let $1 \le k \le 6$. Assume that for the family of probability distributions Θ_k, $1 \le k \le 6$ it exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_{\cdot}^{X|Y}$. Then there exists $1 \le r \le n-1$ such that the RSS experiment without repetition

$$G_{r+1}^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r+1)}} \otimes_{i=r+2}^n P_\theta^X : \theta \in \Theta_k\})$$

is not exhaustive for the RSS experiment

$$G_r^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
We prove the assertion by a contradiction argument. Assume that $\forall\, r \in \{1, ..., n-1\}$ the experiment G_{r+1}^n is exhaustive for G_r^n, i.e. there exist Markov kernels $L^r : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}) \to [0,1]$, $1 \le r \le n-1$, such that

$$\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X(B) = \int_{\mathbb{R}^n} L^r(x_1,\ldots,x_n,B) \otimes_{i=1}^{r+1} P_\theta^{X_{[i]}^{(r+1)}}(dx_i) \otimes_{i=r+2}^n P_\theta^X(dx_i)$$

for every $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$. Define $L : \mathbb{R}^n \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ by

$$L(\cdot, B) := \int_{\mathbb{R}^{n(n-2)}} L^1(x_1^1,\ldots,x_n^1,B) \prod_{i=2}^{n-2} L^i(x_1^i,\ldots,x_n^i,dx_1^{i-1},\ldots,dx_n^{i-1}) L^{n-1}(\cdot, dx_1^{n-2},\ldots,dx_n^{n-2}).$$

Obviously, L is a Markov kernel and it therefore has to satisfy

$$
\begin{aligned}
& \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{i=1}^{n} P_\theta^{X_{[i]}^{(n)}}(dx_i) \\
= & \int_{\mathbb{R}^{n(n-1)}} L^1(x_1^1, \ldots, x_n^1, B) \prod_{i=2}^{n-2} L^i(x_1^i, \ldots, x_n^i, dx_1^{i-1}, \ldots, dx_n^{i-1}) \\
& L^{n-1}(\cdot, dx_1^{n-2}, \ldots, dx_n^{n-2}) \prod_{i=1}^{n} P_\theta^{X_{[i]}^{(n)}}(dx_i) \\
= & \int_{\mathbb{R}^{n(n-2)}} L^1(x_1^1, \ldots, x_n^1, B) \prod_{i=2}^{n-2} L^i(x_1^i, \ldots, x_n^i, dx_1^{i-1}, \ldots, dx_n^{i-1}) \\
& \prod_{i=1}^{n-1} P_\theta^{X_{[i]}^{(n-1)}}(dx_i^{n-2}) P_\theta^X(dx_n^{n-2}) \\
= & \int_{\mathbb{R}^n} L^1(x_1^1, \ldots, x_n^1, B) \prod_{i=1}^{2} P_\theta^{X_{[i]}^{(2)}}(dx_i^1) \prod_{i=3}^{n} P_\theta^X(dx_i^1)
\end{aligned}
$$

which, by the assumptions made, is equal to $\otimes_{i=1}^n P_\theta^X(B)$. This would imply that the Markov kernel serves for the exhaustivity relation

$$
\otimes_{i=1}^n P_\theta^X(B) = \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, B) \prod_{i=1}^{n} P_\theta^{X_{[i]}^{(n)}}(dx_i)
$$

for every $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$ which leads to a contradiction, since we have proved in Proposition 67 that the statistical experiment G_n^n is not exhaustive for the statistical experiment G_1^n.

□

By applying Definition 15 and Theorem 16 in the Prerequisites chapter, we show that the order relation between the RSS experiments does not hold: $G_1^n \subseteq G_n^n$, i.e. the common RSS experiment is not more informative than the SRS experiment of size n. The mentioned theorem can be applied only in the case of dominated experiments, i.e. it has to exist a σ-finite measure ν such that $\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} \ll \nu$ for every $\theta \in \Theta_k$. Therefore, for this to happen, we restrict to the parameter spaces Θ_5 and Θ_6 respectively.

Theorem 69 Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$, $k \in \{5, 6\}$ there exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Then for the RSS experiments without repetition

$$
G_r^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n : \theta \in \Theta_k\}), \quad 1 \le r \le n, k \in \{5, 6\},
$$

it holds:

1. G_n^n is not more informative than G_1^n (the common RSS experiment is not more informative than the SRS experiment of size n).

2. The family of statistical experiments G_r^n, $1 \leq r \leq n$ is not informative.

Proof.
1. By Proposition 67 we have that the RSS experiment G_n^n is not exhaustive for the SRS experiment of size n, therefore the assertion follows by applying Theorem 16.
2. By Proposition 68 it follows that $\exists\, r \in \{1, \ldots, n-1\}$ such that the experiment G_{r+1}^n is not more informative than the experiment G_r^n. Therefore, the assertion follows analogously by applying Theorem 16.

□

Corollary 70 For the parameter spaces Θ_5 and Θ_6, there exists (D, W) decision problem with W a continuous loss function such that $G_1^n \overset{(D,W)}{\not\subseteq} G_n^n$. If D is a locally compact metric space then for this decision problem it follows that $\exists\ \rho_1 \in \mathcal{R}(G_1^n, D)$ such that for every $\rho_n \in \mathcal{R}(G_n^n, D)$

$$W_\theta \rho_1(\otimes_{i=1}^n (P_\theta^X)) < W_\theta \rho_n(\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}), \quad \theta \in \Theta_5, \Theta_6.$$

Proof. The assertion follows by the definition of the informational order 12 and by Theorem 69.

□

4.3 Examples

Assume in a fixed statistical experiment $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ with a decision problem (D, W) and a functional $f(\theta)$, the problem is to find an estimator k which minimizes the risk $\int W_\theta(k(x)) P_\theta(dx)$ simultaneously for all $\theta \in \Theta$. Formulated in this way, the problem is from an applied point of view not meaningful, it might happen that an estimator minimizes the risk at one $\theta \in \Theta$ but we obtain different minimizing estimators for the rest of the family. The situation is different if we restrict for examples to mean unbiased estimators. The definition of *more informative* statistical experiments has also this impediment. Although one can construct experiments which are more informative via the randomization criterion (for example via sufficient statistics), for a large class of experiments, including therefore our family of RSS experiments, the informational order does not take place. Even though, this is the first step to be done when comparing statistical experiments. An important break in this direction occurred when LeCam[19] stated instead of the question when a statistical experiment is more informative than other, another question, how much do we loose respectively win if we use the one experiment instead of another. The answer to this question can be given by using the definition of Δ-*deficiency*.[6] which can be applied for every comparison of statistical experiments. The calculus of the Δ-deficiency between the RSS experiment without repetition and the SRS starts by considering for example first

$$\delta(RSS, SRS) = \delta(G_n^n, G_1^n)$$

$$= \inf_{L \in \mathcal{K}(G_n^n, \mathcal{B}(\mathbb{R})^n)} \sup_{\theta \in \Theta} \left\| L \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} - \otimes_{i=1}^n P_\theta^X \right\|$$

$$= \inf_{L \in \mathcal{K}(G_n^n, \mathcal{B}(\mathbb{R})^n)} \sup_{\theta \in \Theta} \sup_{|f(x)| \leq 1} \left| \int_{\mathbb{R}^n} f(x) d(L \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} - \otimes_{i=1}^n P_\theta^X) \right|$$

$$= \inf_{L \in \mathcal{K}(G_n^n, \mathcal{B}(\mathbb{R})^n)} \sup_{\theta \in \Theta} \sup_{|f(x)| \leq 1} \left| \int \int f(x) L(y, dx) \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(dy) - \int f(x) \otimes_{i=1}^n P_\theta^X(dx) \right|$$

where we have denoted by $\mathcal{K}(G_n^n, \mathcal{B}(\mathbb{R})^n)$ the set of all Markov kernels from G_n^n to $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$, $L \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}$ is then a probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n)$ defined by

$$L \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(B) = \int_{\mathbb{R}^n} L(x, B) \otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(B)(dx), \quad B \in \mathcal{B}(\mathbb{R})^n$$

and the norm is treated as the total variation in $L_\infty(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n))$ (see JACOD[18], section V.4a). It follows directly from the definition of the deficiency that $\delta(RSS, SRS) = 0 \Leftrightarrow$ the RSS experiment G_n^n is more informative than the SRS experiment G_1^n, which by Theorem 69 is not the case. This considerations for the deficiency between the two relevant experiments are topic for further research.

Let us restrict now, as in the case of a fixed experiment, only to the case of mean unbiased estimators for regular functionals and try to see how the relationship between RSS and SRS looks like for this particular class of decisions.

Definition 71 A decision problem (Θ, D, W) is called an *estimation problem* if there is a function $f : \Theta \to D$ such that W_θ depends on θ only through $f, \theta \in \Theta$.

Definition 72 A function $f : \Theta \to \mathbb{R}^k$ admits an *mean unbiased estimator* k for the statistical experiment $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ if k satisfies $\int k P_\theta = f(\theta)$, $\theta \in \Theta$. The set of all mean unbiased estimates of f for the statistical experiment mentioned is denoted with $H(\mathcal{E}, f)$.

Remark 73 Let $\mathcal{E} = (E, \mathcal{B}, \{P_\theta : \theta \in \Theta\})$ be a statistical experiment and (Θ, D, W) an estimation problem. If k is an estimator then it exists a randomized decision function given

[6] Let $\mathcal{E}$ and $\mathcal{F}$ be two statistical experiments. Then the experiment $\mathcal{E}$ is ϵ-deficient with respect to the experiment $\mathcal{F}$ if for every decision problem (Θ, D, W) with a bounded continuous loss function and for every $\beta_2 \in \mathcal{B}(\mathcal{F}, D)$ there is $\beta_1 \in \mathcal{B}(\mathcal{E}, D)$ such that $\beta_1(W_\theta, P_\theta) \leq \beta_2(W_\theta, Q_\theta) + \epsilon \|W_\theta\|$, $\theta \in \Theta$. We denote $\mathcal{F} \overset{\epsilon}{\subseteq} \mathcal{E}$. Define $\delta(E, F) = inf\{\epsilon > 0 : \mathcal{F} \overset{\epsilon}{\subseteq} \mathcal{E}\}$. Then $\Delta(E, F) := max\{\delta(\mathcal{E}, \mathcal{F}), \delta(\mathcal{F}, E)\}$ is called the *deficiency* between $\mathcal{E}$ and $\mathcal{F}$.

by $\rho(\cdot, B) = 1_B \circ k$ and the risk function for this decision function reduces to $W_\theta \rho P_\theta = \int W_\theta(k(x)) P_\theta(dx), \quad \theta \in \Theta$.

For the purposes of this section, elements of $L_2(G_r^n) := \bigcap_{\theta \in \Theta} L_2(\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X)$ are called estimates for the experiment G_r^n for every $1 \le r \le n$ and for particular choices of Θ.

Estimating Regular Statistical Functionals of Degree 1

Definition 74 Let $\{P_\theta : \theta \in \Theta\}$ a family of probability distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then the functional $f : \Theta \to \mathbb{R}$ is called a regular statistical functional if it exists a measurable function $h : \mathbb{R}^m \to \mathbb{R}$ such that $f(\theta) = \int_{R^m} h(x_1, \ldots, x_m) \prod_{i=1}^m P_\theta^X(dx_i)$ for all $\theta \in \Theta$. The function h is called the kernel of the functional and the smallest m such that the equality above holds, is called the degree of the kernel h, respectively the degree of the functional f.

Despite the fact that most of the following considerations hold for arbitrary parameter spaces $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$, where Θ_0 is defined in equation (21), the relevant assertions will hold only for parameter space for which in the SRS experiment we can construct minimum variance unbiased estimators, for example parameter spaces for which the order statistic is sufficient and complete. This is possible for example for the parameter space defined in equation (27): $\Theta_6 := \{P^{XY} \in \mathcal{MB}(\mathbb{R})^2 : P^{XY} \ll \lambda^2, P^Y \in \Delta_2, P^X \in \Delta_2\}$ and $\Delta_2 = \{P \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : \int h_i(x_1, \ldots, x_m) \prod_{i=1}^m P(dx) < \infty, i = 1, \ldots, l\})$.

Proposition 75 Let the SRS experiment be $G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n : \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_6\}$ and let $h : \mathbb{R}^m \to \mathbb{R}$ be a symmetric kernel of degree m. The the U-statistic $U_n(x_1, \ldots, x_n) := E.[h(X_1, \ldots, X_m) \mid O_n(x_1, \ldots, x_n)] = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} h(x_{i_1}, \ldots, x_{i_m})$, $n \ge m$, is an unbiased estimator for $f(\theta) = \int_{R^m} h(x_1, \ldots, x_m) \prod_{i=1}^m P_\theta^X(dx_i)$, $\theta \in \Theta_6$, i.e. $U_n \in H(G_1^n, f)$. Moreover, it minimizes the convex risk among all mean unbiased estimators for $f(\theta)$.

Proof.
See for example Theorem 3.2.7. in Pfanzagl [21].

□

The question is how the estimator U_n behaves if chosen as an estimator for all the other RSS experiments without repetition. Since we have proved that the RSS experiment is not more informative than the SRS experiment, there exists no Markov kernel to consistently construct from SRS estimators better RSS estimators, which is possible in the case of the existence of sufficient statistics, for example. The choice of the same U_n as an estimator in the RSS experiment is based on considerations for the unbiasedness in the RSS experiment for the chosen functional. The next proposition, typical for the RSS problematic, is used in many papers dealing with RSS.

Proposition 76 If $m = 1$ then $U_n \in H(G_r^n, f)$ for every f regular statistical functional of degree 1 for the SRS experiment, $f(\theta) = \int h(x) P_\theta^X(dx)$, $\theta \in \Theta_k$, $\Theta_k \in \Theta_0$, $k \in \mathbb{N}$.

Proof.
By direct use of the fundamental equation (11):

$$E_{G_r^n}[U_n(x_1,\ldots,x_n)] = \int U(x_1,\ldots,x_n) \prod_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i)$$

$$\begin{aligned}
&= \int \frac{1}{n}\sum_{i=1}^{n} h(x_i) \prod_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i) \\
&= \frac{1}{n}\sum_{i=1}^{r} \int h(x) P_\theta^{X_{[i]}^{(r)}}(dx) + \frac{n-r}{n}\int h(x)P_\theta^X(dx) \\
&= \frac{r}{n}\int h(x)\frac{1}{r}\sum_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}}(dx) + \frac{n-r}{n}\int h(x)P_\theta^X(dx) \\
&= \frac{r}{n}\int h(x)P_\theta^X(dx) + \frac{n-r}{n}\int h(x)P_\theta^X(dx) = f(\theta)
\end{aligned}$$

for every $\theta \in \Theta_k$.

□

We show that for a particular choice of the parameter space, the converse of the previous result is also true.

Proposition 77 Assume we are in the case of perfect RSS experiment, i.e.

$$G_n^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^{X_{(i:n)}} : \theta \in \tilde{\Theta}_k\})$$

$\tilde{\Theta}_k \in \Theta_0$, $k \in \mathbb{N}$. Let f a regular statistical functional of degree m for the SRS experiment, $f(\theta) = \int h(x)P_\theta^X(dx)$, $\theta \in \tilde{\Theta}_k$. Then $m = 1 \Leftrightarrow U_n \in H(G_n, f)$.

Proof.
$(U_n \in H(G_n^n, f)) \Rightarrow (m = 1)$:
Let $\theta \in \tilde{\Theta}_k$. By direct use of the fundamental equation (11) it follows that:

$$\begin{aligned}
f(\theta) &= \int_{\mathbb{R}^m} h(x_1,\ldots,x_m) \prod_{i=1}^{m} P_\theta^X(dx_i) \\
&= \int_{\mathbb{R}^m} h(x_1,\ldots,x_m) \prod_{i=1}^{m} \left(\frac{1}{n}\sum_{j=1}^{n} P_\theta^{X_{(j:n)}}(dx_i)\right) \\
&= \int_{\mathbb{R}^m} h(x_1,\ldots,x_m) \frac{1}{n^m} \sum_{j_1,\ldots,j_m}^{n} P_\theta^{X_{(j_1:n)}}(dx_1)\ldots P_\theta^{X_{(j_m:n)}}(dx_m) \\
&= \frac{1}{n^m} \sum_{j_1,\ldots,j_m}^{n} \int_{\mathbb{R}^m} h(x_1,\ldots,x_m) P_\theta^{X_{(j_1:n)}}(dx_1)\ldots P_\theta^{X_{(j_m:n)}}(dx_m)
\end{aligned}$$

$$= \int_{\mathbb{R}^n} \frac{1}{n^m} \sum_{j_1,\dots,j_m}^{n} h(x_{j_1},\dots,x_{j_m}) \prod_{i=1}^{n} P_\theta^{X_{(i:n)}}(dx_i)$$

$$= \int_{\mathbb{R}^n} V_n(x_1,\dots,x_n) \prod_{i=1}^{n} P_\theta^{X_{(i:n)}}(dx_i)$$

where we have denoted by V_n the V-statistic $V_n(x_1,\dots,x_n) = \frac{1}{n^m}\sum_{j_1,\dots,j_m}^n h(x_{j_1},\dots,x_{j_m})$ Suppose further that $U_n \in H(G_n^I, f))$, i.e. $f(\theta) = \int_{\mathbb{R}^n} U_n(x_1,\dots,x_n)\prod_{i=1}^n P_\theta^{X_{(i:n)}}(dx_i)$ for every $\theta \in \tilde{\Theta}_k$. Therefore we have that $\int_{\mathbb{R}^n}(U_n(x_1,\dots,x_n) - V_n(x_1,\dots,x_n))\prod_{i=1}^n P_\theta^{X_{(i:n)}}(dx_i) = 0$ for every $\theta \in \tilde{\Theta}_k$. By Lemma 31 and Proposition 60 it follows that $U_n = V_n \otimes_{i=1}^n P_\theta^{X_{(i:n)}} - a.s$ which is possible only in the case $m = 1$.

$(m = 1) \Rightarrow U_n \in H(G_n^n, f))$:
Particular case of the Proposition 76. □

Let us now consider arbitrary convex loss functions for evaluating estimators. A loss function is convex if $t \mapsto W_\theta(t)$ is convex. A natural convex loss function is $W_\theta(t) :=\mid t - f(\theta) \mid$, a mathematically more convenient one is the quadratic loss function $W_\theta(t) := (t - f(\theta))^2$. More generally, $W_\theta(t) := C(t - f(\theta))$ is a convex loss function if C is a function attaining its minimum at 0.

Proposition 78 For any convex loss function which depends only of a regular statistical functional f of degree 1, the relation holds:

$$\int_{\mathbb{R}^n} W_\theta(U_n(x_1,\dots,x_n)) \prod_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i) \le \int_{\mathbb{R}} W_\theta(h(x)) P_\theta^X(dx),$$

where U_n is the U-Statistic of degree 1 and $\theta \in \Theta_k$, $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$.

Proof.
By the convexity of the loss function and making use of the fundamental equation (11) it follows:

$$\int_{\mathbb{R}^n} W_\theta(U_n(x_1,\dots,x_n) \prod_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i) =$$

$$= \int_{\mathbb{R}^n} W_\theta\left(\frac{1}{n}\sum_{i=1}^n h(x_i)\right) P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i)$$

$$\le \int_{\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n W_\theta h(x_i) P_\theta^{X_{[i]}^{(r)}}(dx_i) \prod_{i=r+1}^{n} P_\theta^X(dx_i)$$

$$\frac{r}{n}\int W_\theta(h(x))P_\theta^X(x) + \frac{n-r}{n}\int W_\theta(h(x))P_\theta^X(x)$$

$$= \int W_\theta(h(x))P_\theta^X(x)$$

for every $\theta \in \Theta_k$. □

Therefore for a decision problem $(\mathbb{R}, W)$ where we only know that W is a convex loss function we can only derive an order relation between the experiment $(\mathbb{R}, \mathcal{B}, \{P_\theta^X : \theta \in \Theta_k\})$ and G_r^n, the second one being more informative then the first one. This comparison shows that the U-Statistic for a regular statistical functional of degree 1, based on a sample of RSS random variables will have always a smaller risk given any convex loss function, then the U-Statistic based on a single observation from the target unknown distribution, P_θ^X. Further, we consider the more specific quadratic loss function and recall in our framework, the perhaps most relevant result for the applications dealing with RSS.

Proposition 79 Consider the common quadratic loss function $W_\theta = (t - f(\theta))^2$, $\theta \in \Theta_6$ where $f(\theta)$ is a regular statistical functional of degree 1, $f(\theta) = P_\theta^X(dx)$. Then the following inequality holds:

$$\mathrm{var}_{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X}[U_n(x_1, \ldots, x_n)] \leq \mathrm{var}_{\otimes_{i=1}^n P_\theta^X}[k(x_1, \ldots, x_n)] \tag{35}$$

for every $\theta \in \Theta_6$, $k \in H(G_1^n, f)$ and U_n the U-statistic of degree 1.

Proof.
Let $\theta \in \Theta_6$. Let $k \in H(G_1^I, f)$ and denote by $P_\theta^{Z[i]} := P_\theta^{X_{[i]}^{(r)}}$ for all $1 \leq i \leq r$ and $P_\theta^{Z[i]} := P_\theta^X$ for all $r+1 \leq i \leq n$. Then by the theorem of Lehmann-Scheffe and by the fundamental equation (11) it follows that $\int_{\mathbb{R}^n} (k(x_1, \ldots, x_n) - f(\theta))^2 \prod_{i=1}^n P_\theta^X(dx_i) \geq$

$$\begin{aligned}
&\geq \int_{\mathbb{R}^n} (U_n(x_1, \ldots, x_n) - f(\theta))^2 \prod_{i=1}^n P_\theta^X(dx_i) \\
&= \frac{1}{n} \int_{\mathbb{R}} (h(x) - f(\theta))^2 P_\theta^X(dx) \\
&= \frac{1}{n^2} \sum_{i=1}^n \int_{\mathbb{R}} h(x)^2 P_\theta^{Z_{[i]}}(dx) - \frac{1}{n} f(\theta)^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \left(\int_{\mathbb{R}} (h(x) - f(\theta))^2 P_\theta^{Z_{[i]}}(dx) + \int_{\mathbb{R}} h(x) P_\theta^{Z_{[i]}}(dx) \right) - \\
&- \frac{1}{n} \left(\frac{1}{n} \sum_{i=1}^n \int h(x) P_\theta^{Z_{[i]}}(dx) \right)^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \int_{\mathbb{R}} (h(x) - f(\theta))^2 P_\theta^{Z_{[i]}}(dx) + \frac{1}{n^2} \sum_{i=1}^n \left(\int_{\mathbb{R}} (h(x) - f(\theta)) P_\theta^{Z_{[i]}}(dx) \right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathbb{R}^n}(h(x_i)-f(\theta))^2\prod_{i=1}^{r}P_\theta^{X_{[i]}^{(r)}}(dx_i)\prod_{i=r+1}^{n}P_\theta^X(dx_i)+\\
&+ \frac{1}{n^2}\sum_{i=1}^{n}\left(\int_{\mathbb{R}}(h(x)-f(\theta))P_\theta^{Z_{[i]}}(dx)\right)^2\\
&= \int_{\mathbb{R}^n}W_\theta(U_n(x_1,\ldots,x_n))\prod_{i=1}^{r}P_\theta^{X_{[i]}^{(r)}}(dx_i)\prod_{i=r+1}^{n}P_\theta^X(dx_i)+\\
&+ \frac{1}{n^2}\sum_{i=1}^{r}\left(\int_{\mathbb{R}}(h(x)-f(\theta))P_\theta^{X_{[i]}^{(r)}}(dx)\right)^2
\end{aligned}
$$

Since the second term is positive the assertion follows immediately. □

Since we have restricted to the class of unbiased estimators, the last proposition cannot imply the assertion $G_1^n \overset{(D,W)}{\subseteq} G_r^n, \quad 2 \leq r \leq n$ for the decision problem $(\mathbb{R},(t-f(\theta)^2)$ where $f(\theta)=\int h(x)P_\theta^X(dx)$ is a regular statistical functional of degree 1. Therefore, the decision problem $(\mathbb{R},(t-f(\theta)^2)$ cannot serve as an example to lead to the result "the SRS experiment is not more informative that the RSS experiment" that would imply the non-comparability of the RSS experiments with the SRS by the informational order definition. Even though, we affirm that the result is strong enough to use it from an applied point of view, analogously to a result like an estimator is the minimum variance unbiased estimator in the class of all unbiased estimators in a fixed statistical experiment.

The next proposition serves for a motivation of the treatment of the comparison of the RSS experiment with the SRS experiment via Markov kernels. We show that there exists no sufficient statistic that could induce the RSS experiment such that the sufficient statistic would lead to the sample mean as estimator of the expectation. Therefore the next natural step is to search not for a sufficient statistic but for a Markov kernel which does not has to exist necessary only through the existence of a sufficient or exhaustive statistic. Since we need the existence of the second moments with respect to X, we restrict to the parameter space Θ_6. Denote by $\hat{\Theta}_k \subset \Theta_k$ the subset $\hat{\Theta}_k := \{\theta \in \Theta_k : \exists i \in \{1,\ldots,r\}, P_\theta^{X_{[i]}^{(r)}} \neq P_\theta^X\}$.

Theorem 80 For every $k:\mathbb{R}^n \to \mathbb{R}$, $\otimes_{i=1}^n P_\theta^X$-unbiased estimator of $f(\theta)$, regular statistical functional of degree 1 it holds

$$U_n(x_1,\ldots,x_n) \neq E_{\otimes_{i=1}^n P_\theta^X}[k \mid S=(x_1,\ldots,x_n)], \otimes_{i=1}^n P_\theta^X S^{-1} - a.s.$$

$\forall\theta \in \hat{\Theta}_6$, where U_n is the U-statistic of degree 1 and $S:\mathbb{R}^n \to \mathbb{R}^n$ is assumed to be a sufficient statistic for the SRS experiment $G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_6\})$ such that $\otimes_{i=1}^n P_\theta^X \circ S^{-1} = \otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X$ for every $\theta \in \Theta_6$.

Proof.
For the parameter space Θ_6, the order statistic O_n is a minimal sufficient statistic for the

SRS experiment, $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_6\})$. Let h be the kernel of the functional to be estimated, i.e. $f(\theta) = \int h(x) P_\theta^X(dx)$. Then $h(x)$ is an unbiased estimator of $f(\theta)$. By the minimal sufficiency of the order statistic, the conditional expectation of h given O_n, denote it by $g_O(x_1, \ldots, x_n) = E_{\otimes_{i=1}^n P_\theta^X}[h \mid O_n = (x_1, \ldots, x_n)]$, is independent of $\theta \in \Theta_6$. By the Lehmann-Scheffe theorem it follows that the estimator $g_O \circ O_n = U_n$ is the minimum variance unbiased estimator in the SRS experiment, i.e.

$$\text{var}_{\otimes_{i=1}^n P_\theta^X}[g_O(O_n(x_1, \ldots, x_n)] \leq \text{var}_{\otimes_{i=1}^n P_\theta^X}[k(x_1, \ldots, x_n)], \quad \theta \in \Theta_6$$

for any k, $\otimes_{i=1}^n P_\theta^X$-unbiased estimator of $f(\theta)$. Consider now the sufficient statistic S. Since $\otimes_{i=1}^n P_\theta^X \circ O_n^{-1} \neq \otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X$ for every $\theta \in \Theta_6$ we notice that $S \neq O_n$ $\otimes_{i=1}^n P_\theta^X$-a.s. Let k be an arbitrary $\otimes_{i=1}^n P_\theta^X$-unbiased estimator of $f(\theta)$. Also by sufficiency of S for the SRS experiment, it follows that the conditional expectation of k given S, $g_S(x_1, \ldots, x_n) = E_{\otimes_{i=1}^n P_\theta^X}[k \mid S = (x_1, \ldots, x_n)]$ is independent of $\theta \in \Theta_6$. By the Rao-Blackwell theorem, the measurable function $g_S \circ S$ is an unbiased estimator of $f(\theta)$, therefore, since $g_O \circ O_n$ is minimum variance unbiased estimator we obtain

$$\begin{aligned} & \text{var}_{\otimes_{i=1}^n P_\theta^X}[g_O(O_n(x_1, \ldots, x_n)] \leq \text{var}_{\otimes_{i=1}^n P_\theta^X}[g_S(S(x_1, \ldots, x_n)] \\ \Leftrightarrow \quad & \text{var}_{\otimes_{i=1}^n P_\theta^X}[g_O(O_n(x_1, \ldots, x_n)] \leq \text{var}_{\otimes_{i=1}^n P_\theta^X S^{-1}}[g_S(x_1, \ldots, x_n)] \end{aligned}$$

for every $\theta \in \Theta_6$. Let $\theta \in \hat{\Theta}_6 \subset \Theta_6$. Therefore $\exists i \in \{1, \ldots, r\}$ such that $P_\theta^{X_{[i]}^{(p)}} \neq P_\theta^X$. Moreover, we know from Proposition 79 that

$$\begin{aligned} & \text{var}_{\otimes_{i=1}^n P_\theta^X S^{-1}}[U_n(x_1, \ldots, x_n)] + \frac{1}{n^2} \sum_{i=1}^r \left(\int_{\mathbb{R}} (h(x) - f(\theta)) P_\theta^{X_{[i]}^{(r)}}(dx) \right)^2 \leq \\ \leq \quad & \text{var}_{\otimes_{i=1}^n P_\theta^X}[U_n(x_1, \ldots, x_n)]. \end{aligned}$$

When restricting to the parameter space $\hat{\Theta}_6$ the stipulated inequality is strict and it follows that

$$\begin{aligned} & \text{var}_{\otimes_{i=1}^n P_\theta^X S^{-1}}[U_n(x_1, \ldots, x_n)] < \text{var}_{\otimes_{i=1}^n P_\theta^X}[g_O(O_n(x_1, \ldots, x_n)] \\ \Rightarrow \quad & \text{var}_{\otimes_{i=1}^n P_\theta^X S^{-1}}[U_n(x_1, \ldots, x_n)] < \text{var}_{\otimes_{i=1}^n P_\theta^X S^{-1}}[g_S(x_1, \ldots, x_n)]. \end{aligned}$$

Therefore, $U_n \neq g_S$, $\otimes_{i=1}^n P_\theta^X S^{-1}$-a.s. $\forall \theta \in \hat{\Theta}_k$.

□

Takahasi [27] proved in the case of perfect RSS, i.e. when $X = Y$ P-a.s., that the variance of the RSS estimator, $\text{var}_{\otimes_{i=1}^n P_\theta^{X(i:n)}}[\frac{1}{n} \sum_{i=1}^n x_i]$ decreases as n increases. Consider Θ_6 as given by equation (27). We generalize Takahasi's proof by showing that the assertion holds uniformly in the family of RSS experiments

$$G_r^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^n P_\theta^X : \theta \in \Theta_6\})$$

for the estimation of a regular statistical functional of degree 1.

Proposition 81 Consider the quadratic loss function $W_\theta = (t - f(\theta))^2$, $\theta \in \Theta_6$ where $f(\theta)$ is a regular statistical functional of degree 1. Then the following inequality holds:

$$\operatorname{var}_{\otimes_{i=1}^{r+1} P_\theta^{X_{[i]}^{(r+1)}} \otimes_{i=r+2}^{n} P_\theta^X}[U_n(x_1, \ldots, x_n)] \leq \operatorname{var}_{\otimes_{i=1}^{r} P_\theta^{X_{[i]}^{(r)}} \otimes_{i=r+1}^{n} P_\theta^X}[U_n(x_1, \ldots, x_n)]$$

for every $1 \leq r \leq n$, $\theta \in \Theta_6$ and U_n the U-statistic of degree 1.

Proof.
The inequality above is equivalent to:

$$\operatorname{var}\left[\frac{1}{n}\left(\sum_{i=1}^{r} h(X_{[i]}^{(r)}) + \sum_{i=r}^{n} h(X_i)\right)\right] - \operatorname{var}\left[\frac{1}{n}\left(\sum_{i=1}^{r+1} h(X_{[i]}^{(r+1)}) + \sum_{i=r+2}^{n} h(X_i)\right)\right] \geq 0$$

$$\Leftrightarrow \quad \frac{1}{n^2}\sum_{i=1}^{r} \operatorname{var}[h(X_{[i]}^{(r)})] - \frac{1}{n^2}\sum_{i=1}^{r+1} \operatorname{var} h(X_{[i]}^{(r+1)}) + \frac{\operatorname{var}[h(X)]}{n^2} \geq 0$$

Denote by $f_{(i:r)}(x)$ the density function of the i-th order statistic in a sample of size r and by $f_{X|Y}(x)$ the conditional density function of X given Y. Then the following recurrence relation holds:

$$f_{(i:r)}(x) = \frac{r+1-i}{r+1} f_{(i:r+1)}(x) + \frac{i}{r+1} f_{(i+1:r+1)}(x)$$

for every $r \in \mathbb{N}$, $1 \leq i \leq r$ and $x \in \mathbb{R}$. According to this recurrence relation, we can express the k-th moments of the RSS random variables recurrently:

$$\begin{aligned}
E[h(X_{[i]}^{(r)})]^k &= \int h(x)^k P_\theta^{X_{[i]}^{(r)}}(dx) \\
&= \int\int h(x)^k f_{X|Y=y}(x) f_{(i:r)}(y)\,dx\,dy \\
&= \int\int h(x)^k f_{X|Y=y}(x)\left(\frac{r+1-i}{r+1} f_{(i:r+1)}(y) + \frac{i}{r+1} f_{(i+1:r+1)}(y)\right)dx\,dy \\
&= \frac{r+1-i}{r+1} E[h(X_{[i]}^{(r+1)})]^k + \frac{i}{r+1} E[h(X_{[i+1]}^{(r+1)})]^k
\end{aligned}$$

Therefore

$$\begin{aligned}
\operatorname{var}[h(X_{[i]}^{(r)})] &= E[h(X_{[i]}^{(r)})]^2 - [Eh(X_{[i]}^{(r)})]^2 \\
&= \left(\frac{r+1-i}{r+1}\operatorname{var}[h(X_{[i]}^{(r+1)})] + \frac{i}{r+1}\operatorname{var}[h(X_{[i+1]}^{(r+1)})]\right) + \\
&+ \left(\frac{r+1-i}{r+1} E[h(X_{[i]}^{(r+1)})] + \frac{i}{r+1} E[h(X_{[i+1]}^{(r+1)})]\right)^2 + \\
&+ \left(\frac{r+1-i}{r+1} E[h(X_{[i]}^{(r+1)})]^2 + \frac{i}{r+1} E[h(X_{[i+1]}^{(r+1)})]^2\right).
\end{aligned}$$

Using the recurrence relation we obtain:

$$\sum_{i=1}^{r} \text{var}[h(X_{[i]}^{(r)})] = \frac{r}{r+1}\sum_{i=1}^{r+1} \text{var}[h(X_{[i]}^{(r+1)})] + \frac{r}{r+1}\sum_{i=1}^{r+1} E[h(X_{[i]}^{(r+1)})]^2 - $$
$$- \frac{1}{(r+1)^2}\sum_{i=1}^{r} \Big((r+1-i)E[h(X_{[i]}^{(r+1)})] + iE[h(X_{[i+1]}^{(r+1)})]\Big)^2 .$$

Additionally, remark that given euquation (11) the following equation holds

$$\text{var}[h(X)] = \frac{1}{r+1}\sum_{i=1}^{r+1} \text{var}[h(X_{[i]}^{(r+1)})] + \frac{1}{r+1}\sum_{i=1}^{r+1}(E[h(X_{[i]}^{(r+1)})] - E[h(X)])^2. \quad (36)$$

Summarizing, we obtain for the left hand side of equation (36) multiplied by n^2:

$$\sum_{i=1}^{r} \text{var} h(X_{[i]}^{(r)}) - \sum_{i=1}^{r+1} \text{var} h(X_{[i]}^{(r+1)}) + \text{var} h(X) =$$
$$= \sum_{i=1}^{r} \text{var} h(X_{[i]}^{(r)}) - \sum_{i=1}^{r+1} \text{var} h(X_{[i]}^{(r+1)}) + \frac{1}{r+1}\sum_{i=1}^{r+1} \text{var}[h(X_{[i]}^{(r+1)})] +$$
$$+ \frac{1}{r+1}\sum_{i=1}^{r+1}(E[h(X_{[i]}^{(r+1)})] - E[h(X)])^2$$
$$= \frac{r}{r+1}\sum_{i=1}^{r+1} E[h(X_{[i]}^{(r+1)})]^2 - \frac{1}{(r+1)^2}\sum_{i=1}^{r} \Big((r+1-i)E[h(X_{[i]}^{(r+1)})] + iE[h(X_{[i+1]}^{(r+1)})]\Big)^2 +$$
$$+ \frac{1}{r+1}\sum_{i=1}^{r+1}(E[h(X_{[i]}^{(r+1)})] - E[h(X)])^2$$
$$= \frac{1}{(r+1)^2}\Bigg(r(r+1)\sum_{i=1}^{r+1} E[h(X_{[i]}^{(r+1)})]^2 - \sum_{i=1}^{r} \Big((r+1-i)E[h(X_{[i]}^{(r+1)})]+$$
$$+ iE[h(X_{[i+1]}^{(r+1)})]\Big)^2\Bigg) + \frac{1}{r+1}\sum_{i=1}^{r+1}(E[h(X_{[i]}^{(r+1)})] - E[h(X)])^2$$
$$= \frac{1}{(r+1)^2}\sum_{i=1}^{r} i(r+i-1)\Big(E[h(X_{[i+1]}^{(r+1)})] - E[h(X_{[i]}^{(r+1)})]\Big)^2 +$$
$$+ \frac{1}{r+1}\sum_{i=1}^{r+1}(E[h(X_{[i]}^{(r+1)})] - E[h(X)])^2 \geq 0.$$

□

In order to prove that a possible information order for estimation problems in the family of RSS experiments is possible when restricting to unbiased estimators, we need to prove first that the sample mean is the minimum variance unbiased estimator in every fixed RSS experiment in the family.

Proposition 82 For every $1 \leq r \leq n$ fixed but arbitrary, the sample mean is the minimum variance unbiased estimator in the perfect RSS experiment

$$G_r^{n,perfect} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^r P_\theta^{X(i:r)} \otimes_{i=r+1}^n P_\theta^X : \theta \in \Theta_6\})$$

Proof.
Define the measurable function $S : (\mathbb{R}^n \mathcal{B}(\mathbb{R})^n) \rightarrow (\mathbb{R}^n \mathcal{B}(\mathbb{R})^n)$,

$$(x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_r, O_{n-r}(x_{r+1}, \ldots, x_n))$$

where O_{n-r} is the ordinary order statistic function. The function S is then a sufficient and complete statistic for the experiment $G_r^{n,perfect}$. To prove the sufficiency let $B \in \mathcal{B}(\mathbb{R})^n$ such that $B = B_1 \times \cdots \times B_n$, B_i right open interval in $\mathbb{R}$, $1 \leq i \leq n$ and $\theta \in \Theta_6$. Then the conditional expectation with respect to the probability distribution in $G_r^{n,perfect}$ of 1_B given S is

$$\begin{aligned}
& E_\theta[1_B \mid (x_1, \ldots, x_r, O_{n-r}(x_{r+1}, \ldots, x_n))] \\
= \; & E_\theta[1_{B_1 \times \cdots \times B_r} 1_{B_{r+1} \times \cdots \times B_n} \mid (x_1, \ldots, x_r, O_{n-r}(x_{r+1}, \ldots, x_n))] \\
= \; & E_\theta[1_{B_1 \times \cdots \times B_r} \mid (x_1, \ldots, x_r)] E_\theta[1_{B_{r+1} \times \cdots \times B_n} \mid O(x_{r+1}, \ldots, x_n)] \\
= \; & 1_{B_1 \times \cdots \times B_r} \frac{1}{(n-r)!} \sum_{\pi \in \mathcal{S}_{n-r}} 1_{\tau_\pi (B_{r+1} \times \cdots \times B_n)}, \quad \otimes_{i=1}^r P_\theta^{X(i:r)} \otimes_{i=r+1}^n P_\theta^X - a.s.
\end{aligned}$$

Denote by Γ^n the system of all finite unions of figures in $\mathbb{R}^n$. Then the function $H : \mathbb{R}^n \times \Gamma^n \rightarrow [0, 1]$

$$H(\cdot, B) = 1_{B_1 \times \cdots \times B_r} \frac{1}{(n-r)!} \sum_{\pi \in \mathcal{S}_{n-r}} 1_{\tau_\pi (B_{r+1} \times \cdots \times B_n)}$$

can be extended $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$ to a probability measure on $\mathcal{B}(\mathbb{R})^n$, say $\tilde{H}$ such that $\tilde{H}$ is a version of the conditional distribution given S and independent of $\theta \in \Theta_6$, $\otimes_{i=1}^r P_\theta^{X(i:r)} \otimes_{i=r+1}^n P_\theta^X - a.s.$, therefore, S is a sufficient statistic.

For the completeness, let $h \in \bigcap_{\theta \in \Theta_6} L_1(G_r^I)$ such that

$$\int h(x_1, \ldots, x_r, O(x_{r+1}, \ldots, x_n)) \otimes_{i=1}^r P_\theta^{X(i:r)} \otimes_{i=r+1}^n P_\theta^X (dx_1, \ldots, dx_n) = 0 \quad \theta \in \Theta_6.$$

By the transformation formula it follows that,

$$\int h(x_1, \ldots, x_n) \prod_{i=1}^r P_\theta^{X(i:r)}(dx_i) P_\theta^{O_{n-r} X^{n-r}} (dx_{r+1}, \ldots, dx_n) = 0, \quad \theta \in \Theta_6.$$

Since the order statistic is complete, by the Proposition 60 and the Proposition 31 it follows that $h = 0$, $\otimes_{i=1}^r P_\theta^{X(i:r)} P_\theta^{O_{n-r} X^{n-r}}$-a.s., which implies that S is a complete statistic for the

RSS experiment $G_r^{n,perfect}$. Let $\frac{1}{n}\sum_{i=1}^r h(x_i) + \frac{n-r}{n}h(x_{r+1})$ be an unbiased estimator of $f(\theta)$. Let $g(y_1,\ldots,y_n) := E_\theta(\frac{1}{n}\sum_{i=1}^r h(x_i) + \frac{n-r}{n}h(x_{r+1}) \mid S(x_1,\ldots,x_n) = (y_1,\ldots,y_n))$, $\otimes_{i=1}^r P_\theta^{X_{(i:r)}} \otimes_{i=r+1}^n P_\theta^X - a.s.$ Therefore, by the theorem of Lehmann-Scheffe, the estimator $g \circ S$ is an unbiased estimator and has the smallest variance among all unbiased estimators with respect to the G_r^n experiment, i.e.

$$\mathrm{var}_\theta[g(S(x_1,\ldots,x_n)] \leq \mathrm{var}_\theta[k(x_1,\ldots,x_n]$$

for any k unbiased estimator of $f(\theta)$ and for each $\theta \in \Theta_6$. The uniformly minimum variance estimator for the functional $f(\theta)$ is then:

$$\begin{aligned} g(S(x_1,\ldots,x_n)) &= E_\theta(\frac{1}{n}\sum_{i=1}^r h(x_i) + \frac{n-r}{n}h(x_{r+1}) \mid S(x_1,\ldots,x_n)) \\ &= E_\theta(\frac{1}{n}\sum_{i=1}^r h(x_i) \mid (x_1,\ldots,x_r,O_{n-r}(x_{r+1},\ldots,x_n)) + \\ &+ \frac{n-r}{n}E_\theta(x_{r+1} \mid (x_1,\ldots,x_r,O_{n-r}(x_{r+1},\ldots,x_n)) \\ &= \frac{1}{n}\sum_{i=1}^r h(x_i) + \frac{n-r}{n}\frac{1}{n-r}\sum_{i=r+1}^n h(x_i) \\ &= \frac{1}{n}\sum_{i=1}^n h(x_i) \\ &= U_n(x_1,\ldots,x_n). \end{aligned}$$

□

Corollary 83 Let $f(\theta) = \int h(x)P_\theta^X(dx)$, $\theta \in \Theta_6$ a functional to be estimated. Then the following inequality holds in the family of perfect RSS experiments:

$$\mathrm{var}_{\otimes_{i=1}^{r+1} P_\theta^{X_{(i:r+1)}} \otimes_{i=r+2}^n P_\theta^X}[U_n(x_1,\ldots,x_n)] \leq \mathrm{var}_{\otimes_{i=1}^{r+1} P_\theta^{X_{(i:r+1)}} \otimes_{i=r+2}^n P_\theta^X}[k(x_1,\ldots,x_n)], \quad \theta \in \Theta_6$$

for every $1 \leq r \leq n-1$. Here U_n is the U-statistic for estimating $f(\theta)$ and k is any unbiased estimator in the $G_r^{n,perfect}$ experiment.

Proof.
Let $1 \leq r \leq n-1$ and $\theta \in \Theta_6$. By the Proposition 81 it follows that

$$\begin{aligned} &\mathrm{var}_{\otimes_{i=1}^{r+1} P_\theta^{X_{(i:r+1)}} \otimes_{i=r+2}^n P_\theta^X}[U_n(x_1,\ldots,x_n)] \leq \mathrm{var}_{\otimes_{i=1}^{r} P_\theta^{X_{(i:r)}} \otimes_{i=r+1}^n P_\theta^X}[U_n(x_1,\ldots,x_n)] \leq \\ \leq\ &\mathrm{var}_{\otimes_{i=1}^{r+1} P_\theta^{X_{(i:r+1)}} \otimes_{i=r+2}^n P_\theta^X}[k(x_1,\ldots,x_n)] \end{aligned}$$

where the last inequality follows from the minimum variance property of the U-Statistic in the experiment $G_r^{n,perfect}$. □

5 RSS Experiment with Repetition and SRS of Smaller Dimension

5.1 The Informational Order

In the last section we have treated comparisons of RSS experiments without repetition on probability spaces of the same dimension. We concluded that the family of RSS experiments without repetition is not more informative, i.e. there exists no Markov kernels which should recursively generate the randomization of the experiments. In this section we are interested in the comparison of RSS experiments with repetition with the SRS experiment of a smaller dimension. We conclude that in this case the RSS with repetition is more informative than the SRS experiment. The results are particular relevant for further comparison of the RSS experiments. At the end of the chapter we present an example to motivate the results also from an applied point of view.

Recall the RSS experiments with repetition for the parameter spaces $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$:

$$G_{t_1}^{t_1} \otimes \cdots \otimes G_{t_n}^{t_n} := (\mathbb{R}^{\sum_{i=1}^n t_i}, \mathcal{B}(\mathbb{R})^{\sum_{i=1}^n t_i}, \{\otimes_{j=1}^{t_1} P_\theta^{X_{[j]}^{(t_1)}} \otimes \cdots \otimes_{j=1}^{t_n} P_\theta^{X_{[n]}^{(t_p)}} : \theta \in \Theta_k\}),$$

$t_i \in \mathbb{N}$, $1 \leq i \leq n$. The parameter space Θ_0 is defined in equation (21).

Proposition 84 Let $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$ an arbitrary parameter space. The experiment $(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^{X_{[i]}^{(n)}} : \theta \in \Theta_k\})$ is exhaustive for the experiment $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \{P_\theta^X : \theta \in \Theta_k\})$, $n \in \mathbb{N}$.

Proof.
Let $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$. Then, by the fundamental equation (11) it follows that

$$\begin{aligned}
P_\theta^X(A) &= \frac{1}{n}\sum_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(A) \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} 1_A(x) P_\theta^{X_{[i]}^{(n)}}(dx) \\
&= \int_{\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n 1_A(x_i) \prod_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(dx_i) \\
&= \int_{\mathbb{R}^n} \frac{1}{n}\sum_{j=1}^n 1_{\tau_{(1j)}(A \times \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n-1})}(x_1, \ldots, x_n) \prod_{i=1}^n P_\theta^{X_{[i]}^{(n)}}(dx_i)
\end{aligned}$$

Therefore, the Markov kernel denoted by $K^n : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}) \to [0, 1]$,

$$K^n(\cdot, A) = \frac{1}{n}\sum_{j=1}^n 1_{\tau_{(1j)}(A \times \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n-1})}, \quad A \in \mathcal{B}(\mathbb{R}) \tag{37}$$

fulfills the exhaustivity equation.

□

Remark 85 Let $\theta_0 \in \Theta_k$, $k \in \mathbb{N}$ such that $P_{\theta_0}^{XY} = P_{\theta_0}^{X} P_{\theta_0}^{Y}$, i.e. X and Y are independent. Then $P_{\theta_0}^{X_{[i]}^{(n)}} = P_{\theta_0}^{X}$ for every $1 \leq i \leq n$. Therefore, for such a distribution the Markov kernel defined in (37) serves for the exhaustivity equation $P_{\theta_0}^{X}(A) = \frac{1}{n}\sum_{i=1}^{n} P_{\theta_0}^{X}(A)$, for every θ_0 which stands for independence and for every $A \in \mathcal{B}(\mathbb{R})$. In the case of dominated experiments, this implies that a SRS experiment of size n is always more informative than a SRS experiment of size 1. In other words, this serves for a proof that every decision made with more than one observation in an experiment with i.i.d. observations, is better than a decision made upon one observation.

Proposition 86 Assume that for the family of probability distributions Θ_k, $1 \leq k \leq 6$ it exists a version of the conditional distribution $P_{\theta}^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_{\theta}^{X} - a.s.$. Denote it by $P_{\cdot}^{X|Y}$. Then the kernel K^n defined in equation (37) is unique in the following sense: for every other kernel $L : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}) \to [0,1]$ which fulfills the exhaustivity equation $P_{\theta}^{X}(A) = \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, A) \prod_{i=1}^{n} P_{\theta}^{X_{[i]}^{(n)}}(dx_i)$ for every $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$ it holds:

$$\int_{\mathbb{R}^n} L(x_1, \ldots, x_n, A) \prod_{i=1}^{n} P_{\cdot}^{X|Y=y_i} = \int_{\mathbb{R}^n} K^n(x_1, \ldots, x_n) \prod_{i=1}^{n} P_{\cdot}^{X|Y=y_i}$$

$\otimes_{i=1}^{n} P_{\theta}^{Y_{(i:n)}} - a.s.$ and for every $A \in \mathcal{B}(\mathbb{R})$.

Proof.
Assume there exists a Markov kernel $L : \mathbb{R}^n \times \mathcal{B}(\mathbb{R}) \to [0,1]$ such that

$$P_{\theta}^{X}(A) = \int_{\mathbb{R}^n} L(x_1, \ldots, x_n, A) \prod_{i=1}^{n} P_{\theta}^{X_{[i]}^{(n)}}(dx_i)$$

for every $A \in \mathcal{B}(\mathbb{R})$ and $\theta \in \Theta_k$. Therefore

$$\int_{\mathbb{R}^n} (L(x_1, \ldots, x_n, A) - K^n(x_1, \ldots, x_n, A)) \prod_{i=1}^{n} P_{\theta}^{X_{[i]}^{(n)}}(dx_i) = 0$$

$$\Leftrightarrow \quad \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} (L(x_1, \ldots, x_n, A) - K^n(x_1, \ldots, x_n, A)) \prod_{i=1}^{n} P_{\cdot}^{X|Y=y}(dx_i) \prod_{i=1}^{n} P_{\theta}^{Y_{(i:n)}}(dy_i) = 0$$

Since for the families Θ_k the order statistic is complete, the assertion follows then by Lemma 31 and Proposition 60. □

Lemma 87 There exists a unique Markov kernel $Q^{t_1,t_2} : \mathbb{R}^{\sum_{i=1}^2 t_i} \times \mathcal{B}(\mathbb{R})^2 \to [0,1]$ such that

$$Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) = \prod_{i=1}^{2} K^{t_i}(x_1^i, \dots, x_{t_i}^i; B_i) \tag{38}$$

$\forall B = B_1 \times B_2 \; B_i \in \mathcal{B}(\mathbb{R}), 1 \le i \le 2$. For each $B \in \mathcal{B}(R)^2$ it holds $Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) =$

$$\int K^{t_2}(x_1^2, \dots, x_{t_2}^2; B_{z_1}) K^{t_1}(x_1^1, \dots, x_{t_1}^1; dz_1) = \int K^{t_1}(x_1^1, \dots, x_{t_1}^1; B_{z_2}) K^{t_2}(x_1^2, \dots, x_{t_2}^2; dz_2)$$

where $B_{z_1} := \{z_2 : (z_1, z_2) \in B\}$ (respectively $B_{z_2} := \{z_1 : (z_1, z_2) \in B\}$).

Proof.
That Q^{t_1,t_2} defined by $Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) = \int K^{t_2}(x_1^2, \dots, x_{t_2}^2; B_{z_1}) K^{t_1}(x_1^1, \dots, x_{t_1}^1; dz_1)$ is a probability measure on $\mathcal{B}(\mathbb{R})^2$ is obvious. Analogously, Q^{t_1,t_2} defined by $Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) = \int K^{t_1}(x_1^1, \dots, x_{t_1}^1; B_{z_2}) K^{t_2}(x_1^2, \dots, x_{t_2}^2; dz_2)$ is a probability measure on $\mathcal{B}(\mathbb{R})^2$. The uniqueness follows by arguments similar like for the uniqueness of the product measure, see for example Bauer [3], section 21. Let us check the property (38). Denote for each $x \in \mathbb{R}^{t_2}$ and $B \in \mathcal{B}(\mathbb{R})^2$ by $s_B(x, \cdot)$ the $\mathcal{B}(\mathbb{R})$ measurable function $z_1 \mapsto K^{t_2}(x, B_{z_1})$ on $\mathbb{R}$. Therefore, $Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) = \int s_B(x_1^2, \dots, x_{t_2}^2, z_1) K^{t_1}(x_1^1, \dots, x_{t_1}^1; dz_1)$. Let $B \in \mathcal{B}(\mathbb{R})^2$ such that $B = B_1 \times B_2$, $B_i \in \mathcal{B}(\mathbb{R})$, $1 \le i \le 2$ and $(x_1^2, \dots, x_{t_2}^2) \in \mathbb{R}^{t_2}$. Then

$$\begin{aligned} s_{B_1 \times B_2}(x_1^2, \dots, x_{t_2}^2, z_1) &= K^{t_2}(x_1^2, \dots, x_{t_2}^2, (B_1 \times B_2)_{z_1}) \\ &= \frac{1}{t_2} \sum_{j=1}^{t_2} 1_{\tau_{1j}\{(B_1 \times B_2)_{z_1} \times \mathbb{R} \cdots \times \mathbb{R}\}}(x_1, \dots, x_{t_2}) \\ &= K^{t_2}(x_1^2, \dots, x_{t_2}^2, B_2) 1_{B_1}(z_1) \end{aligned}$$

and therefore

$$\begin{aligned} Q^{t_1,t_2}(x_j^i : 1 \le i \le 2, 1 \le j \le t_i; B) &= \int K^{t_2}(x_1^2, \dots, x_{t_2}^2, B_2) 1_{B_1}(z_1) K^{t_1}(x_1^1, \dots, x_{t_1}^1; dz_1) \\ &= K^{t_2}(x_1^2, \dots, x_{t_2}^2; B_2) K^{t_1}(x_1^1, \dots, x_{t_1}^1; B_1) \end{aligned}$$

for every $B \in \mathcal{B}(\mathbb{R})^2$ such that $B = B_1 \times B_2$.

□

Lemma 88 There exists a unique Markov kernel $Q^{t_1,\dots,t_n} : \mathbb{R}^{\sum_{i=1}^n t_i} \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ for $n \ge 2$ such that that

$$Q^{t_1,\dots,t_n}(x_j^i : 1 \le i \le n, 1 \le j \le t_i; B) = \prod_{i=1}^{n} K^{t_i}(x_1^i, \dots, x_{t_i}^i; B_i) \tag{39}$$

$\forall B = B_1 \times \cdots \times B_n \; B_i \in \mathcal{B}(\mathbb{R}), 1 \le i \le n$.

Proof.
If $n = 2$ than the existence and uniqueness are proved by the previous lemma. Assume now that such a Markov kernel exists for $n-1$, i.e. it exists $Q^{t_1,\dots,t_{n-1}} : \mathbb{R}^{\sum_{i=1}^{n-1} t_i} \times \mathcal{B}(\mathbb{R})^{n-1} \to [0,1]$ such that that

$$Q^{t_1,\dots,t_{n-1}}(x_j^i : 1 \le i \le n-1, 1 \le j \le t_i; B) = \prod_{i=1}^{n-1} K^{t_i}(x_1^i, \dots, x_{t_i}^i; B_i) \tag{40}$$

$\forall B = B_1 \times \cdots \times B_{n-1}$ $B_i \in \mathcal{B}(\mathbb{R}), 1 \le i \le n-1$. We define $Q^{t_1,\dots,t_n} : \mathbb{R}^{\sum_{i=1}^{n} t_i} \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ by $Q^{t_1,\dots,t_n}(x_j^i : 1 \le i \le n, 1 \le j \le t_i; B) :=$

$$\int Q^{t_1,\dots,t_{n-1}}(x_j^i : 1 \le i \le n-1, 1 \le j \le t_i; B_{z_n}) K^{t_n}(x_1^{t_n}, \dots, x_{t_n}^{t_n}; dz_n)$$

for every $B \in \mathcal{B}(\mathbb{R})^n$ and where $B_{z_n} := \{(z_1, \dots, z_{n-1}) : (z_1, \dots, z_n) \in B\}$. Then, analogously as in the previous lemma, the Markov kernel satisfies equation (39). The uniqueness follows by the same arguments as in Bauer [3], Proposition 22.3.

□

Theorem 89 Let $\Theta_k \subseteq \Theta_0$, $k \ge 0$ be an arbitrary parameter space. Let $n \in \mathbb{N}$ be fixed but arbitrary. The RSS experiment with repetition

$$G_{t_1}^{t_1} \otimes \cdots \otimes G_{t_n}^{t_n} = (\mathbb{R}^{\sum_{i=1}^n t_i}, \mathcal{B}(\mathbb{R})^{\sum_{i=1}^n t_i}, \{\otimes_{j=1}^{t_1} P_\theta^{X_{[j]}^{(t_1)}} \otimes \dots \otimes_{j=1}^{t_n} P_\theta^{X_{[j]}^{(t_n)}} : \theta \in \Theta_k\}),$$

$t_i \in N^*$, $1 \le i \le n$, is exhaustive for the SRS experiment

$$G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
Let $Q^{t_1,\dots,t_n} : \mathbb{R}^{\sum_{i=1}^n t_i} \times \mathcal{B}(\mathbb{R})^n \to [0,1]$ be the Markov kernel such that

$$Q^{t_1,\dots,t_n}(x_j^i : 1 \le i \le n, 1 \le j \le t_i; B) = \prod_{i=1}^{n} K^{t_i}(x_1^i, \dots, x_{t_i}^i; B_i) \tag{41}$$

$\forall B = B_1 \times \cdots \times B_n$ $B_i \in \mathcal{B}(\mathbb{R}), 1 \le i \le n$. Then we affirm that

$$\otimes_{i=1}^n P_\theta^X(B) = \int Q^{t_1,\dots,t_n}(x_j^i : 1 \le i \le n, 1 \le j \le t_i; B) \prod_{i=1}^{n} \prod_{j=1}^{t_i} P_\theta^{X_{[j]}^{(t_i)}}(dx_j^i)$$

for all $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$. We prove the assertion by making use of the previous lemma and several use of the proposition of Fubini. Additionally, recall the Markov kernel K^{t_i}

defined in equation (37) and serving for the exhaustivity in Proposition 84. Let $B \in \mathcal{B}(\mathbb{R})^n$ and $\theta \in \Theta_k$. Then

$$\begin{aligned}
&\int Q^{t_1,\dots,t_n}(x^i_j : 1 \le i \le n, 1 \le j \le t_i; B) \prod_{i=1}^{n}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j) \\
= &\int Q^{t_1,\dots,t_{n-1}}(x^i_j : 1 \le i \le n-1, 1 \le j \le t_i; B_{z_n}) \\
&K^{t_n}(x^{t_n}_1, \dots, x^{t_n}_{t_n}; dz_n) \prod_{i=1}^{n}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j) \\
= &\int Q^{t_1,\dots,t_{n-1}}(x^i_j : 1 \le i \le n-1, 1 \le j \le t_i; B_{z_n}) \\
&P^X_\theta(dz_n) \prod_{i=1}^{n-1}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j) \\
= &\int Q^{t_1,\dots,t_{n-2}}(x^i_j : 1 \le i \le n-2, 1 \le j \le t_i; B_{z_n,z_{n-1}}) \\
&K^{t_{n-1}}(x^{t_n}_1, \dots, x^{t_n}_{t_n}; dz_{n-1}) P^X_\theta(dz_n) \prod_{i=1}^{n-2}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j) \\
= &\int Q^{t_1,\dots,t_{n-2}}(x^i_j : 1 \le i \le n-2, 1 \le j \le t_i; B_{z_n,z_{n-1}}) \\
&P^X_\theta(dz_{n-1}) P^X_\theta(dz_n) \prod_{i=1}^{n-2}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j).
\end{aligned}$$

Here we define for an iterative use the sets $B_{z_n,\dots,z_k} \in \mathcal{B}(\mathbb{R})^{k-1}$,

$$B_{z_n,\dots,z_k} := \{(z_1, \dots, z_{k-1}) : (z_1, \dots, z_k) \in B_{z_n,\dots,z_k}\}, \quad 2 \le k \le n.$$

By repeating the steps above we obtain
$\int Q^{t_1,\dots,t_n}(x^i_j : 1 \le i \le n, 1 \le j \le t_i; B) \prod_{i=1}^{n}\prod_{j=1}^{t_i} P_\theta^{X^{(t_i)}_{[j]}}(dx^i_j) =$

$$\begin{aligned}
&\int K^{t_1}(x^{t_1}_1, \dots, x^{t_1}_{t_1}; B_{z_n,\dots,z_2}) \otimes_{i=2}^{n} P^X_\theta(dz_i) \prod_{j=1}^{t_1} P_\theta^{X^{(t_1)}_{[j]}}(dx^i_j) \\
= &\int P^X_\theta(B_{z_n,\dots,z_2}) \prod_{i=1}^{n} P^X_\theta(dz_i) \\
= &\otimes_{i=1}^{n} P^X_\theta(B).
\end{aligned}$$

□

Theorem 90 Assume $\Theta_k \subseteq \Theta_0$, $k \geq 0$ are dominated families of probability distributions. Then for the RSS experiment with repetition

$$G_{t_1}^{t_1} \otimes \cdots \otimes G_{t_n}^{t_n} := (\mathbb{R}^{\sum_{i=1}^n t_i}, \mathcal{B}(\mathbb{R})^{\sum_{i=1}^n t_i}, \{\otimes_{i=1}^n \otimes_{j=1}^{t_1} P_\theta^{X_{[j]}^{(t_i)}} : \theta \in \Theta_k\}), \quad t_i \in \mathbb{N}, 1 \leq i \leq n$$

and the SRS experiment

$$G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\})$$

the informational order holds

$$G_1^n \subseteq G_{t_1}^{t_1} \otimes \cdots \otimes G_{t_n}^{t_n}.$$

Proof.
By the last proposition it follows that the experiment $G_{t_1}^{t_1} \otimes \cdots \otimes G_{t_n}^{t_n}$ is exhaustive for the experiment G_n. Therefore, the assertion follows by the Definition 15 and the Theorem 16.

□

Corollary 91 For every decision problem (Θ_k, D, W) with a continuous loss function and $\forall\ \rho_1 \in \mathcal{R}(G_1^n, D)$, $\exists\ \rho_2 \in \mathcal{R}(\otimes_{i=1}^n G_{t_i}^{t_i}, D)$ such that $W_\theta \rho_2 \otimes_{i=1}^n \otimes_{j=1}^{t_i} P_\theta^{X_{[j]}^{(t_i)}} \leq W_\theta \rho_1 \otimes_{i=1}^n P_\theta^X, \quad \theta \in \Theta.$

Proof.
By the definition of the informational order.

□

From an applied point of view, this result is of relative importance. In the case of estimation, the last corollary implies the comparison of an estimator based on $X_1, \ldots, X_n$, i.i.d P_θ^X observations with an estimator based on $X_{[j]}^{(t_i)} \sim P_\theta^{X_{[j]}^{(t_i)}}$, $1 \leq i \leq n$, $1 \leq j \leq t_i$ and all independent. Thus and estimator based on n observations is compared to an estimator based on $\sum_{i=1}^n t_i$ observations. In the common estimation problems, the precision of an estimator in terms of variance increases with the sample size, i.e. the variance of an estimator decreases as the sample size increases. Despite this, there are a few examples where within the SRS experiment this assertion does not hold, i.e. where an estimator with less observations will be more precise. When comparing the SRS experiment, G_1^n with the RSS experiment $\otimes_{i=1}^n G_{t_i}^{t_i}$, this impediment will not happen. Let us give some examples, the first being the analogous example from the end of the chapter 4.

5.2 Examples

Example 1.
Consider the balanced RSS experiment with repetition, i.e. $t_1 = \ldots = t_n = p$, $p \in \mathbb{N}$ for the parameter space $\Theta_k \subseteq \Theta_0$, $k \in \mathbb{N}$

$$\otimes_{i=1}^n G_p^p = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1} \otimes_{j=1}^p P_\theta^{X_{[j]}^{(p)}} : \theta \in \Theta_k\})$$

and the SRS of size n

$$G_1^n = (\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_k\}).$$

Let the decision problem be an estimation problem. Assume we want to estimate $f(\theta) = \int h(x)P_\theta^X(dx)$, $\theta \in \Theta_k$ where h is the integrable kernel of the regular functional of degree 1. The loss function is the common quadratic loss. The informational order between the RSS experiments with repetition and the SRS of smaller dimension implies that for every estimator we choose in the SRS experiment, there is a better estimator in the RSS experiment. We know, for certain parameter spaces, the U-statistic of degree 1 is the minimum variance unbiased estimator within the SRS experiment, therefore, we choose U-statistic $U_n = n^{-1}\sum_{i=1}^n h(x_i)$ to estimate the functional above. We define a new estimator $k : \mathbb{R}^{np} \to \mathbb{R}$ by making use of the Markov kernel defined in equation (41), $k(x_j^i : 1 \le j \le p, 1 \le i \le n) :=$

$$\int_{\mathbb{R}^n} U_n(y_1, \dots, y_n)Q^{p,\dots,p}(x_j^i : 1 \le j \le p, 1 \le i \le n; dy_1, \dots, dy_n).$$

Proposition 92

$$k(x_j^i : 1 \le j \le p, 1 \le i \le n) = \frac{1}{np}\sum_{i=1}^n \sum_{j=1}^p h(x_j^i).$$

Proof.

$$\begin{aligned}
& \int_{\mathbb{R}^n} U_n(y_1, \dots, y_n)Q^{p,\dots,p}(x_j^i : 1 \le i \le n, 1 \le j \le p;\ dy_1, \dots, dy_n) \\
= & \int_{\mathbb{R}^n} \frac{1}{n}\sum_{i=1}^n h(y_i)\prod_{i=1}^n K^p(x_1^i, \dots, x_p^i, dy_i) \\
= & \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^n} h(y_i)K^p(x_1^i, \dots, x_p^i, dy_i) \\
= & \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}} h(y_i)\left(\frac{1}{p}\sum_{j=1}^p 1_{\tau_{1j}(dy_j \times \mathbb{R} \dots \times \mathbb{R})}(x_1^i, \dots, x_p^i)\right) \\
= & \frac{1}{np}\sum_{i=1}^n \sum_{j=1}^p \int_{\mathbb{R}} h(y_i)\delta_{\{\mathrm{pr}_1 \tau_{1j}(x_1^i, \dots, x_p^i)\}}(dy_i) = \frac{1}{np}\sum_{i=1}^n \sum_{j=1}^p h(x_j^i)
\end{aligned}$$

□

Therefore, it follows that

$$\operatorname{var}_{\otimes_{i=1}^n \otimes_{j=1}^p P_\theta^{X_{[j]}^{(p)}}}[\frac{1}{np}\sum_{i=1}^n \sum_{j=1}^p h(x_j^i)] \le \operatorname{var}_{\otimes_{i=1}^n P_\theta^X}[\frac{1}{n}\sum_{i=1}^n h(x_i)], \quad \theta \in \Theta_k.$$

Remark also that by integrating with respect to the Markov kernel $Q^{p,\dots,p}$ we obtain the U-statistic as estimator in the RSS experiment. In the case of the common quadratic loss function, the variance of the U-statistic decreases as the sample size increases, so the obtained estimator is here from an applied point of view of relative importance.

Example 2.
To present the second example we need to introduce some more definitions and a proposition relevant to the assertions which we are going to make. For more details on the example within a fixed experiment, see Pfanzagl [21], example 2.7.6.

Definition 93 (Concentration of Estimators) A probability measure Q_1 on $\mathcal{B}(\mathbb{R})$ is more concentrated about 0 than another probability measure Q_2 on $\mathcal{B}(\mathbb{R})$ if

$$Q_1(-t', t'') \geq Q_2(-t', t''), \quad t', t'' \geq 0.$$

Consider we have two real estimators k_1 and k_2 defined with respect to an arbitrary statistical experiment whith the family of probability distributions $\{P_\theta : \theta \in \Theta\}$. A criteria to compare two estimators is their concentration around the parameter to be estimated, say $f(\theta)$. Comparing the concentration of k_1 and k_2 about the parameter $f(\theta)$ can be reformulated as the problem of comparing the concentration of the probability measures $P_\theta \circ (k_i - f(\theta))$ about 0. Consider now o more natural class of loss functions, the class of all subconvex functions.

Definition 94 (Subconvex Loss Functions) A familiy of loss functions $(W_\theta)_{\theta\in\Theta}$, $W_\theta : \mathbb{R}^n \to \mathbb{R}_+$ is called subconvex if $\{t \in \mathbb{R}^n : W_\theta(t) \leq u\}$ is convex for $u \geq 0$ and $\theta \in \Theta$.

Any convex loss functions is subconvex. If $n = 1$, the condition of subconvexity is reduces to the requirement that $t \mapsto W_\theta(t)$ is increasing as t moves away from $f(\theta)$, in each direction. An example of a subconvex loss function which is not convex is for instance, $W_\theta = 1 - 1_{C_{f(\theta)}}$ where $C_{f(\theta)}$ is a measurable convex set containing $f(\theta)$. If two estimators are comparable, simultaneously for all (symmetric) subconvex loss functions, they are comparable with respect to their concentration on all convex sets containing $f(\theta)$ (symmetric about $f(\theta)$). This is the content of the following proposition.

Proposition 95 The following assertions are equivalent.

1. $P_\theta(k_1 \in C) \geq P_\theta(k_2 \in C)$ for every convex set C containing $f(\theta)$ (symmetric about $f(\theta)$).
2. For every subconvex (and symmetric) loss function $(W_\theta)_{\theta\in\Theta}$, the risk of k_1 is smaller than the risk of k_2.

Proof. See Proposition 2.5.3. in Pfanzagl [21]. □

If an estimator, say $k^n : \mathbb{R}^n \to \mathbb{R}$ is defined for every sample size n, one normally expects that the variance of k^{n+1} is smaller than the variance of k^n or that k^{n+1} is more concentrated around the parameter to be estimated than k^n. We want to show to which extent this is true. Consider the family of bivariate normal probability distributions:

$$\Theta_N = \{N_{(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)} : (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2), \mu_x, \mu_y \in \mathbb{R}, \sigma_x^2 > 0, \sigma_y^2 > 0, \rho \in (0, 1)\}.$$

Consider the SRS experiment of dimension n

$$G_1^n = \{\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, \{\otimes_{i=1}^n P_\theta^X : \theta \in \Theta_N\})$$

and the RSS experiment with repetition

$$\otimes_{i=1}^n G_{t_i}^{t_i} = (\mathbb{R}^{\sum_{j=1}^n t_j}, \mathcal{B}(\mathbb{R})^{\sum_{j=1}^n t_j}, \{\otimes_{i=1}^n \otimes_{j=1}^{t_i} P_\theta^{X_{[j]}^{(t_i)}} : \theta \in \Theta_N\}).$$

The problem is to estimate the variance with respect to P_θ^X, i.e. σ_x^2. Let us first treat the problem only within the SRS experiment. Within the SRS experiment, the estimator $s_n^2(x_1, \ldots, x_n) := (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ minimizes the convex risk among all mean unbiased estimators (by the theorem of Lehmann and Sheffe). Moreover, the risk of s_{n+1}^2 is smaller than the risk of s_n^2 for every convex function. We put the question if s_{n+1}^2 is more concentrated than s_n^2. Since $N_{\mu_x,\sigma_x^2}^{n+1}(s_{n+1}^2 \leq \sigma_x^2) \neq N_{\mu_x,\sigma_x^2}^{n}(s_n^2 \leq \sigma_x^2)$, s_{n+1}^2 cannot be more concentrated than s_n^2 on all intervals containing σ_x^2. Therefore, by Proposition 95 it exists at least a subconvex loss function, denote it by L_θ, such that the risk with respect to L_θ of s_n^2 is smaller than the risk with respect to L_θ of s_{n+1}^2. This impediment will not happen if we proceed with the estimation of σ_x^2 within a RSS experiment with repetition where we fix t_i, $1 \leq i \leq n$ such that $\sum_{j=1}^n t_j = n+1$. By Theorem 90, the RSS experiment, $\otimes_{i=1}^n G_{t_i}^{t_i}$ is more informative than the SRS for every decision problem. Therefore, also for $(\mathbb{R}, (L_\theta)_{\theta \in \Theta_k})$. By using the kernel defined in equation (41) we construct a new estimator, denote it by $s_{n+1,RSS}^2$, $s_{n+1,RSS}^2(x_j^i : 1 \leq i \leq n, 1 \leq j \leq t_i) =$

$$\int s_n^2(y_1, \ldots, y_n) Q^{t_1,\ldots,t_n}(x_j^i : 1 \leq i \leq n, 1 \leq j \leq t_i; dy_1, \ldots, dy_n)$$

Therefore, by Corollary 91 it follows that:

$$\begin{aligned} & \int L_\theta(s_{n+1,RSS}^2(x_j^i : 1 \leq i \leq n, 1 \leq j \leq t_i)) \prod_{i=1}^n \prod_{j=1}^{t_i} P_\theta^{X_{[j]}^{(t_i)}}(dx_j^i) \leq \\ \leq & \int L_\theta(s_n^2(x_1, \ldots, x_n)) \prod_{i=1}^n P_\theta^X(dx_i) \\ \leq & \int L_\theta(s_{n+1}^2(x_1, \ldots, x_{n+1})) \prod_{i=1}^{n+1} P_\theta^X(dx_i) \end{aligned}$$

which gives another motivation of the treatment of the comparison between RSS without repetition and the SRS of a smaller dimension.

6 RSS Experiments with Repetition and SRS of the Same Dimension

In this chapter, we investigate the relationship between the RSS experiment with repetition and the SRS of the same dimension. Since the general case has an analog proof, we treat only the balanced case of RSS with repetition, i.e. we consider $t_1 = \ldots = t_n = p$, $p \in \mathbb{N}$. Recall the RSS experiment with repetition for the parameter spaces Θ_k, $1 \leq k \leq 6$ defined in equations (22)-(27).

$$\otimes_{i=1}^{n} G_p^p = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{n} \otimes_{j=1}^{p} P_\theta^{X_{[j]}^{(p)}} : \theta \in \Theta_k\})$$

and the SRS experiment of size np:

$$G_1^{np} = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{np} P_\theta^X : \theta \in \Theta_k\}).$$

Analogously as in chapter 4, we will prove that under certain considerations, the RSS experiment $\otimes_{i=1}^{n} G_p^p$ is not more informative than a SRS experiment of size np.

Proposition 96 Let $1 \leq k \leq 6$ be fixed. Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$ it exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_\cdot^{X|Y}$. Then the RSS experiment with repetition, the balanced case

$$\otimes_{i=1}^{n} G_p^p = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{n} \otimes_{j=1}^{p} P_\theta^{X_{[j]}^{(p)}} : \theta \in \Theta_k\})$$

is not exhaustive for the the SRS experiment of size np:

$$G_1^{np} = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{np} P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
We will prove the proposition by a contradiction argument. Assume that there exists a Markov kernel $L : \mathbb{R}^{np} \times \mathcal{B}(\mathbb{R})^n \to [0, 1]$ such that the required exhaustivity condition is fulfilled, i.e. for every $B \in \mathcal{B}(\mathbb{R})^{np}$ and $\theta \in \Theta_k$,

$$\otimes_{i=1}^{np} P_\theta^X(B) = \int_{\mathbb{R}^{np}} L(x_1^i, \ldots, x_p^i, 1 \leq i \leq n; B) \prod_{i=1}^{n} \prod_{j=1}^{p} P_\theta^{X_{[j]}^{(p)}}(dx_j^i). \tag{42}$$

Making use of the proposition of Fubini, the right-hand side of equation (42) is then equal to

$$\begin{aligned}
&\int_{\mathbb{R}^{np}} \int_{\mathbb{R}^{np}} L(x_1^i, \ldots, x_p^i, 1 \leq i \leq n; B) \prod_{i=1}^{n} \prod_{j=1}^{p} P_\cdot^{X|Y=y_j^i}(dx_j^i) \prod_{i=1}^{n} \prod_{j=1}^{p} P_\theta^{Y_{(j:p)}}(dy_j^i) \\
= &\int_{\mathbb{R}^{np^2}} \int_{\mathbb{R}^{np}} L(x_1^i, \ldots, x_p^i, 1 \leq i \leq n; B) \prod_{i=1}^{n} \prod_{j=1}^{n} P_\cdot^{X|Y=y_{jj}^i}(dx_j^i) \prod_{i=1}^{n} \prod_{j=1}^{p} \prod_{k=1}^{p} P_\theta^{Y_{(j:n)}}(dy_{kj}^i) \\
= &\int_{\mathbb{R}^{np^2}} h_B(y_{jk}^i : 1 \leq i \leq n, 1 \leq k, j \leq p) \prod_{i=1}^{n} \prod_{j=1}^{p} \prod_{k=1}^{p} P_\theta^{Y_{(j:n)}}(dy_{kj}^i)
\end{aligned}$$

where we have denoted for each $B \in \mathcal{B}(\mathbb{R})^{np}$ by h_B the measurable function

$$(y_{kj}^i : 1 \le i \le n, 1 \le k, j \le p) \mapsto \int_{\mathbb{R}^{np}} L(x_1^i, \ldots, x_p^i, 1 \le i \le n; B) \prod_{i=1}^{n} \prod_{j=1}^{p} P_{\cdot}^{X|Y=y_{jj}^i}(dx_j^i).$$

Let $B \in \mathcal{B}(\mathbb{R})^{np}$ such that $B = B_1 \times \ldots \times B_p$, $B_k \in \mathcal{B}(\mathbb{R})^n$, $1 \le k \le p$. We know from Theorem 89 that the experiment $\otimes_{i=1}^n G_p^p$ is exhaustive for the experiment G_1^n and that the Markov kernel defined in equation (41) satisfies

$$(P_\theta^X)^n(B_k) = \int_{\mathbb{R}^{np}} Q^{p,\ldots,p}(x_{k1}^i, \ldots, x_{kp}^i, 1 \le i \le n; B_k) \prod_{i=1}^{n} \prod_{j=1}^{p} P_\theta^{X_{[j]}^{(p)}}(dx_{kj}^i)$$

for every $1 \le k \le p$ and $\theta \in \Theta_k$.
By independence it follows that $\otimes_{i=1}^{np} P_\theta^X(B) = \prod_{k=1}^p (P_\theta^X)^n(B_k) =$

$$\begin{aligned}
&= \prod_{k=1}^{p} \int_{\mathbb{R}^{np}} Q^{p,\ldots,p}(x_{k1}^i, \ldots, x_{kp}^i, 1 \le i \le n; B_k) \prod_{i=1}^{n} \prod_{j,k=1}^{p} P_\theta^{X_{[j]}^{(p)}}(dx_{kj}^i) \\
&= \int_{\mathbb{R}^{np^2}} \prod_{k=1}^{p} Q^{p,\ldots,p}(x_{k1}^i, \ldots, x_{kp}^i, 1 \le i \le n; B_k) \prod_{i=1}^{n} \prod_{j,k=1}^{p} P_\theta^{X_{[j]}^{(p)}}(dx_{kj}^i) \\
&= \int_{\mathbb{R}^{np^2}} \int_{\mathbb{R}^{np^2}} \prod_{k=1}^{p} Q^{p,\ldots,p}(x_{k1}^i, \ldots, x_{kn}^i, 1 \le i \le n; B_k) \\
&\quad \prod_{i=1}^{n} \prod_{j,k=1}^{p} P_{\cdot}^{X|Y=y_{kj}^i}(dx_{kj}^i) \prod_{i=1}^{n} \prod_{j,k=1}^{p} P_\theta^{Y_{(j:p)}}(dy_{kj}^i) \\
&= \int_{\mathbb{R}^{np^2}} g_B(y_{kj}^i : 1 \le i \le n, 1 \le k, j \le p) \prod_{i=1}^{n} \prod_{j,k=1}^{p} P_\theta^{Y_{(j:p)}}(dy_{kj}^i)
\end{aligned}$$

where we have denoted by g_B the measurable function

$$(y_{kj}^i : 1 \le i \le n, 1 \le k, j \le p) \mapsto \int_{\mathbb{R}^{np^2}} \prod_{k=1}^{p} Q^{p,\ldots,p}(x_{k1}^i, \ldots, x_{kn}^i, 1 \le i \le n; B_k)$$
$$\prod_{i=1}^{n} \prod_{j,k=1}^{p} P_{\cdot}^{X|Y=y_{kj}^i}(dx_{kj}^i)$$

for each $B \in \mathcal{B}(\mathbb{R})^{np}$ such that $B = B_1 \times \ldots \times B_p$, $B_k \in \mathcal{B}(\mathbb{R})^n$, $1 \le k \le p$. For each $B \in \mathcal{B}(\mathbb{R})^{np}$, therefore also for our particular choice, the functions $h_B, g_B \in \bigcap_{\theta \in \Theta_k} L_1\left(\otimes_{j=1}^p P_\theta^{Y_{(j:p)}}\right)^{np}$ are by proposition (60) and (31) equal $\left(\otimes_{j=1}^p P_\theta^{Y_{(j:p)}}\right)^{np} - a.s.$. By arguments similar as in in Proposition 67 it follows that the equality of the functions leads to a contradiction argument. $\square$

Theorem 97 Assume that for the family of probability distributions $\{P_\theta^{XY} : \theta \in \Theta_k\}$, $k \in \{5,6\}$ it exists a version of the conditional distribution $P_\theta^{X|Y}$ which is independent of $\theta \in \Theta_k$, $P_\theta^X - a.s.$. Denote it by $P_\cdot^{X|Y}$. Then the RSS experiment with repetition, the balanced case

$$\otimes_{i=1}^n G_p^p = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^n \otimes_{j=1}^p P_\theta^{X_{[j]}^{(p)}} : \theta \in \Theta_k\})$$

is not more informative than the SRS experiment of size np

$$G_1^{np} = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{np} P_\theta^X : \theta \in \Theta_k\}).$$

Proof.
By the previous proposition we have that the RSS experiment with repetition is not more informative than the SRS experiment of the same size. The proof of the theorem follows therefore by Definition 15 and Theorem 16.

□

Corollary 98 For the parameter space Θ_5 or Θ_6 it exists (D,W) decision problem with W a continuous loss function such that $G_1^{np} \overset{(D,W)}{\not\subseteq} \otimes_{i=1}^n G_p^p$. If D is a locally compact metric space then for this decision problem it follows that $\exists\ \rho_1 \in \mathcal{R}(G_1^{np}, D)$ such that for every $\rho_n \in \mathcal{R}(\otimes_{i=1}^n G_p^p, D)$

$$W_\theta \rho_1(\otimes_{i=1}^{np}(P_\theta^X)) < W_\theta \rho_n(\otimes_{i=1}^n \otimes_{j=1}^p P_\theta^{X_{[j]}^{(p)}}), \quad \theta \in \Theta_5(\Theta_6).$$

Proof. By the definition of the informational order. □

7 Appendix
The Ranked Set Sampling and Stratifications

In this chapter we will treat the RSS experiment with repetition, the balanced and perfect case, from another point of view, by regarding the RSS fundamental equation (11) as a particular mixture distribution, or in a statistical terminology, as a particular stratification of the population. The results can be generalized to arbitrary RSS experiments.
Consider the parametrization $\Theta = \{P^X \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : P^X \ll \lambda\}$ and the corresponding perfect RSS experiment with repetition and balanced

$$\otimes_{i=1}^{n} G_p^{p,perfect} = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{n} \otimes_{j=1}^{p} P_\theta^{X_{(j:p)}} : \theta \in \Theta\}), \tag{43}$$

$p \in \mathbb{N}^*$, $1 \leq i \leq n$. Recall also the SRS of size np for the parameter space Θ:

$$G_1^{np} = (\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \{\otimes_{i=1}^{np} P_\theta^X : \theta \in \Theta\}).$$

Let $\theta \in \Theta$. It is shown in Isii [17] and Taga [26] that the stratification of a set of random variables with distribution P_θ^X into p strata may be represented by a decomposition of P_θ^X into functions $\tilde{Q}_\theta^j : \mathbb{R} \to [0, \infty)$, $1 \leq j \leq p$ which, similar to distribution functions, are right-continuous and non-decreasing such that $P_\theta^X(dx) = \sum_{j=1}^{p} \tilde{Q}_\theta^j(x)$ holds for all $x \in \mathbb{R}$. Let $\alpha_j := \lim_{x\to\infty} \tilde{Q}_\theta^j(x)$, for $1 \leq j \leq p$ and define $Q_\theta^j := \alpha_j^{-1}\tilde{Q}_\theta^j$. Then Q_θ^j is the distribution function of the j-th stratum. For a fixed $p \in \mathbb{N}$ nd $\theta \in \Theta$ we define the set of vector functions

$$\mathcal{Q}_\theta := \{(\tilde{Q}_\theta^1, \ldots, (\tilde{Q}_\theta^p) : \sum_{j=1}^{p} \tilde{Q}_\theta^j(x) = P_\theta^X(x), x \in \mathbb{R}\} \tag{44}$$

which will be called *the set of all stratifications with p strata* for P_θ^X. Analogously, if we first fix the weights $\alpha_j := \lim_{x\to\infty} \tilde{Q}_\theta^j(x)$, $1 \leq j \leq p$ then the set

$$\mathcal{Q}_{\theta,\alpha} := \{(\tilde{Q}_\theta^1, \ldots, \tilde{Q}_\theta^p) : \sum_{j=1}^{p} \tilde{Q}_\theta^j(x) = P_\theta^X(x), x \in \mathbb{R}, \alpha_j := \lim_{x\to\infty} \tilde{Q}_\theta^j(x)\}$$

will be called *the set of all stratifications with p strata and fixed weighting vector* for P_θ^X.

Definition 99 (The Stratified Experiment with Fixed Weights) Consider the parameter space $\Theta = \{P^X \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : P^X \ll \lambda\}$. The stratified experiment with p strata and fixed weights is a statistical experiment $(E, \mathcal{B}, \mathcal{P})$ where $E := \mathbb{R}^{\sum_{j=1}^{p} n_j}$, $\mathcal{B} := \mathcal{B}(\mathbb{R})^{\sum_{j=1}^{p} n_j}$, $\mathcal{P} := \{\otimes_{j=1}^{p} \otimes_{i=1}^{n_j} Q_\theta^j : Q_\theta^j := \alpha_j^{-1}\tilde{Q}_\theta^j, (\tilde{Q}_\theta^1, \ldots, \tilde{Q}_\theta^p) \in \mathcal{Q}_{\theta,\alpha}, \theta \in \Theta\}$ where $(n_1, \ldots, n_p)$ is called the sample allocation vector and $(\alpha_1, \ldots, \alpha_p)$ the weighting vector.

The measures $\tilde{Q}_\theta^j$ are absolutely continuous with respect to P_θ^X, i.e. $\tilde{Q}_\theta^j << P_\theta^X$, $1 \le j \le p$. Therefore, there exists the measurable functions $\phi_\theta^j : \mathbb{R} \to [0,1]$ which are the Radon-Nykodim[7] derivatives of $\tilde{Q}_\theta^j$ with respect to P_θ^X, $\phi_\theta^j(x) = \frac{d\tilde{Q}_\theta^j}{dP_\theta^X}(x)$, $P_\theta^X - a.s.$, for which the relation holds $\sum_{j=1}^p \phi_\theta^j(x) = 1$, $P_\theta^X - a.s.$, for every $\theta \in \Theta$. Let two such functions be identified if they are equal except on sets of P_θ^X-measure zero. For each $\theta \in \Theta$ we define the set of vector functions

$$\Phi_\theta := \{(\phi_\theta^1, \dots, \phi_\theta^p) : \phi_\theta^j : \mathbb{R} \to [0,1], \sum_{j=1}^p \phi_\theta^j(x) = 1, P_\theta^X - a.s.\} \tag{45}$$

One can easily see that for each $\theta \in \Theta$ there is a one-to-one correspondence between the sets $\mathcal{Q}_\theta$ and Φ_θ defined in (44) and (45).
Therefore, a stratification for P_θ^X can be identified with a function vector $\phi_\theta \in \Phi_\theta$.
Analogously, there is a ono-to-one correspondence between the sets $\mathcal{Q}_{\theta,\alpha}$ and the set

$$\Phi_{\theta,\alpha} := \{(\phi_\theta^1, \dots, \phi_\theta^p) : \phi_\theta^j : \mathbb{R} \to [0,1], \sum_{j=1}^p \phi_\theta^j(x) = 1, P_\theta^X - a.s., \int \phi_\theta^j(x) P_\theta^X(dx) = \alpha^j\}.$$

Therefore, an alternative expression for a stratified experiment with fixed weights is $(\mathbb{R}^{\sum_{j=1}^p t_j}, \mathcal{B}(\mathbb{R})^{\sum_{j=1}^p n_j}, \mathcal{P})$ where

$$\mathcal{P} := \{\otimes_{j=1}^p \otimes_{i=1}^{n_j} Q_\theta^j : Q_\theta^j := \alpha_j^{-1} \tilde{Q}_\theta^j, d\tilde{Q}_\theta^j = \phi_\theta^j dP_\theta^X, \phi \in \Phi_{\theta,\alpha}, \theta \in \Theta\}.$$

Remark 100 Let $\theta \in \Theta$, $p, n \in \mathbb{N}$ be fixed and arbitrary. Consider the weight vector $\alpha = (p^{-1}, \dots, p^{-1})$ to be fixed.

1. We define $\phi_{SRS} \in \Phi_{\theta,\alpha}$ by $\phi_{SRS} := (p^{-1}, \dots, p^{-1})$ then this stratification vector together with the allocation vector $(n, \dots, n)$ generates the SRS experiment of size np, i.e. the statistical experiment G_1^{np}.

2. We define the stratification vector $\phi_{\theta,RSS} \in \Phi_{\theta,\alpha}$ by

$$\phi_{\theta,RSS}(x) := (p^{-1}\beta_{(1:p)}(P_\theta^X(x)), \dots, p^{-1}\beta_{(p:p)}(P_\theta^X(x))), \quad x \in \mathbb{R}$$

   where the function $\beta_{(i:p)}$ is defined in equation (6). By Proposition 35, we have $P_\theta^{X_{(j:p)}} \ll P_\theta^X$ with $\frac{dP_\theta^{X_{(j:p)}}}{dP_\theta^X}(x) = \beta_{(i:p)}(P_\theta^X(x))$, $\quad P_\theta^X - a.s.$. Therefore the stratification vector ϕ_{RSS} together with the allocation vector $(n, \dots, n)$ generates the RSS experiment with repetition, the perfect and balanced case, i.e. $\otimes_{i=1}^n G_p^{p,perfect}$ defined in equation (43).

[7] Let P and Q be two measures on a σ-field $\mathcal{B}$ in a space E. If P is σ-finite, then Q has a density with respect to P if and only if Q is absolutely continuous with respect to P. The density is called the Radon-Nykodim density of Q with respect to P.

Next, we recall two theorems from Taga [26], relevant for the perspective on the Ranked Set Sampling as a particular stratification of the original probability distribution. We restrict to balanced stratified experiments.

Uniformly Minimum Variance Estimators in Stratified Experiments
Consider $f(\theta) = \int_{\mathbb{R}^m} h(x_1, \ldots, x_m) \prod_{i=1}^m P_\theta^X(dx_i)$, $\theta \in \Theta$, a regular statistical functional of degree m with a symmetric kernel $h \in L_2(\otimes_{i=1}^m P_\theta^X)$. In the balanced stratified case, Taga [26] defines the estimator

$$U_{strat} := \frac{m!}{n^m} \sum_r \frac{1}{r_1! \ldots r_p!} U_r \tag{46}$$

where

$$U_r = \left[\binom{n}{r_1} \cdots \binom{n}{r_p} \right]^{-1} \sum_\tau h(x^1_{\tau(11)}, \ldots, x^1_{\tau(1r_1)}, \ldots, x^p_{\tau(p1)}, \ldots, x^p_{\tau(pr_p)}).$$

Here the first summation should be taken over all combinations $(r_1, \ldots, r_p)$ of non-negative integers such that $\sum_{j=1}^p r_j = m$ and the second summation over all combinations $\tau = (\tau(11), \ldots, \tau(pr_p))$ of positive integers corresponding to each $(r_1, \ldots, r_p)$ s.t. $1 \leq \tau(j1) < \ldots < \tau(jr_j) \leq n$ for $1 \leq j \leq p$.

Theorem 101 (Uniformly Minimum Variance Estimator in Stratified Experiment)
Consider the parameter space $\Theta = \{P^X \in \mathcal{M}(\mathcal{B}(\mathbb{R})) : P^X \ll \lambda\}$ and let $\Phi_{\theta,\alpha}$ be the set of all stratifications with p strata for P_θ^X, $\theta \in \Theta$, corresponding to any fixed weighting vector. Then the estimator U_{strat} is a uniformly minimum variance unbiased estimator for the functional $f(\theta)$ in the stratified balanced experiment:

$$(\mathbb{R}^{np}, \mathcal{B}(\mathbb{R})^{np}, \mathcal{P} := \{\otimes_{j=1}^p \otimes_{i=1}^n Q_\theta^j : Q_\theta^j := \alpha_j^{-1} \tilde{Q}_\theta^j, d\tilde{Q}_\theta^j = \phi_j dP_\theta^X, \phi \in \Phi_{\theta,\alpha}, \theta \in \Theta\}).$$

Proof.
See Theorem 3.1 in Taga [26]. Basically, the proof relies on the sufficiency and completeness of the generalized order statistic, $O_{strat} : \mathbb{R}^{np} \to \mathbb{R}^{np}$,

$$O_{strat}(\{x_1^1, \ldots, x_n^1\}, \ldots, \{x_1^p, \ldots, x_n^p\}) = (O_n(x_1^1, \ldots, x_n^1) \ldots, O_n(x_1^p, \ldots, x_n^p)),$$

where O_n is the usual order statistic of size n.

□

Corollary 102 Fix $\phi \in \Phi_{\theta,\alpha}$, $\phi := \phi_{\theta,RSS}$. Then the estimator U_{strat} is a uniformly minimum variance unbiased estimator for the functional $f(\theta)$ in the Ranked Set Sampling experiment with repetition, the perfect and balanced case in equation (43).

Proof.
Obvious.

□

Optimum Stratification

Suppose now the number of the strata, p, the balanced allocation $(n, \ldots, n)$ are fixed and consider the set of stratifications Φ_θ for a $\theta \in \Theta$. The variance, respective asymptotic variance of the estimator U_{strat} with respect to $\theta \in \Theta$ and under a stratification $\phi \in \Phi_\theta$ is a continuous function of ϕ on the set Φ_θ which is convex and compact with respect to the weak topology (for a proof see Taga [26]). Therefore, there exists an optimum stratification in Φ_θ which attains the minimum of the variance of U_{strat} for all $\theta \in \Theta$ such that the support of P_θ^X contains at least p points. For further details, see Isii [17]. Let us denote by

$$\phi_\theta^o := \arg\min_{\phi \in \Phi_\theta} \lim_{n \to \infty} \mathrm{var}_{\theta,\phi_\theta}[U_{strat}]$$

the stratification vector which attains the minimum of the asymptotic variance.

Theorem 103 Let $\theta \in \Theta$ such that the support of P_θ^X has at least p points. Then asymptotically optimum stratifications ϕ_θ^o exists in Φ_θ. Moreover, in the case $f(\theta) = \int x P_\theta^X(dx)$, ϕ_θ^o coincides with the indicator function vector $(1_{A_j})_{1 \le j \le p}$ of an interval division of the real line, $A_j := [x_{j-1}, x_j)$, $-\infty = x_0 < x_1 < \ldots < x_p = \infty$. Every end point x_j in A_j can be taken at a continuity point of $P_\theta^X(x)$ such that the condition is satisfied simultaneously $x_j = \frac{\mu_j + \mu_{j+1}}{2}$, $1 \le j \le p$, where $\mu_j = \alpha_j^{-1} \int_{A_j} x P_\theta^X(dx)$.

Proof.
See Taga [26].

□

Remark 104 In the case of the estimation of $f(\theta) = \int x P_\theta^X(dx)$, the asymptotic optimal stratisfication coincides with with the optimal stratification, i.e.

$$\arg\min_{\phi \in \Phi_\theta} \lim_{n \to \infty} \mathrm{var}_{\theta,\phi_\theta}[U_{strat}] = \arg\min_{\phi \in \Phi_\theta} \mathrm{var}_{\theta,\phi_\theta}[U_{strat}].$$

Theorem 105 The stratification determined by the RSS experiments, $\phi_{\theta,RSS}$ does not coincide with the optimal stratification of P_θ^X, $\theta \in \Theta$ for the estimation of $f(\theta) = \int x P_\theta^X(dx)$.

Proof.
Since for a $1 \le j \le p$ it exists $y \in \mathbb{R}$ such that $p^{-1}\beta_{(j:p)}(P_\theta^X(y)) \in (0,1)$, there exists no partition $-\infty = x_0 < x_1 < \ldots < x_p = \infty$ of the real line such that

$$p^{-1}\beta_{(j:p)}(P_\theta^X(x)) \equiv 1_{[x_{j-1}, x_j)}$$

simultaneously for all $1 \le j \le p$. Therefore, the assertion follows by Theorem 103. □

Corollary 106 The following inequality holds:

$$\mathrm{var}_{\theta,\phi_\theta^o}[U_{strat}] \le \mathrm{var}_{\theta,\phi_{\theta,RSS}}[U_{strat}] \le \mathrm{var}_{\theta,\phi_{\theta,SRS}^o}[U_{strat}], \quad \theta \in \Theta.$$

Although the Ranked Set Sampling stratification is therefore not the optimal stratification, in terms of the minimization of the variance, it is an compromise one can do in practice, to avoid computationally problems regarding the numerical calculus of the optimal stratification points. Another future research consists in simulations for the calculus of the distance between the optimal stratification ϕ_θ^o and $\phi_{\theta,RSS}$, for diverse, fixed $\theta \in \Theta$.

References

[1] Bai Z., Chen Z. (2003). *On the theory of ranked-set sampling and its ramifications,* (J. Stat. Plann. Inference 109, No.1-2, 81-99)

[2] Barabesi L., Fattorini L. (2002). *Kernel estimators of probability density functions by ranked-set sampling* (Commun. Stat., Theory and Methods 31, No.4, 597-610)

[3] Bauer H. (1978). *Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie, 3.Auflage,* (de Gruyter Lehrbuch)

[4] Bhattacharya P. (1984). *Induced Order Statistics: Theory and Applications,* (Handbook of Statistics 4, 383-403)

[5] Blackwell D. (1951). *Comparison of Experiments,* (Proc. 2nd Berkeley Symp. Math. Stat. Probab., 1, 93-102)

[6] Blackwell D. (1953). *Equivalent comparisons of experiments,* (Ann. Math. Statistics 24, 265-272)

[7] Bohnenblust A., Shapley L., Sherman S. (1949). *Reconnaissance in game theory* (Rand. Research Memorandum 1948/208)

[8] Chen Z. (2000). *On Ranked-Set Sample Quantiles and Their Applications,* (J. Stat. Plann. Inference 83, 125-135)

[9] Chen Z. (2001). *Ranked-Set Sampling with Regression Type Estimators,* (J. Stat. Plann. Inference 92, 181-192)

[10] Cochran W. (1953). *Sampling Techniques,* (Wiley, New York)

[11] David H. (1981). *Order Statistics. 2nd edn.,* (Wiley, New York)

[12] Dell T., Clutter J. (1972). *Ranked set sampling theory with order statistics background,* (Biometrics 28, 545-555)

[13] Efron B. (1979). *Boostrap Methods: another look at the Jackknife,* (Ann. Statist. 7, 1-26)

[14] Halls L., Dell T. (1966). *Trial of ranked-set sampling for forage yields,* (Forest Science, 12, 22-26)

[15] Hettmansperger T. (1995). *The ranked set sample sigh test,* (J. Nonpar. Statist. 4, 263-270)

[16] Heyer H. (1982). *Theory of Statistical Experiments,* (Springer- Berlin-Heidelberg-New York)

[17] Isii, K., Taga Y. (1969). *Optimal stratifications for the multivariate distributions,* (Skand. Aktuar. Tidskr. 24-38)

[18] Jacod J.,Shiryaev A. (1987). *Limit Theorems for Stochastic Processes,* (Springer Verlag, New York etc)

[19] LeCam L. (1964). *Sufficiency and Approximate Sufficiency,* (Ann. Math. Stat., 35, 1419-1455)

[20] McIntyre G. (1952). *A method of unbiased selective sampling, using ranked sets,* (Austral. J. Agric. Res. 3, 385-390)

[21] Pfanzagl J. (1994). *Parametric Statistical Theory,* (de Gruyter Textbook)

[22] Quenouille M. (1949). *Approximate tests of correlation in time series,* (J. Roy. Statist. Soc. Ser. B, 11, 18-84)

[23] Shiryaev A., Spokoiny V. (2000). *Statistical Experiments and Decisions,* (Advanced Series on Statistical Science and Applied Probability, Vol.8)

[24] Stokes S. (1980). *Estimation of the variance using judgement ordered ranked set samples,* (Biometrics 36, 35-42)

[25] Strasser H. (1985). *Mathematical Theory of Statistics,* (de Gruyter Studies in Mathematics)

[26] Taga Y. et all (1967). *Generalized U-Statistics for Stratified Random Samples,* (Skand. Aktuar. Tiskr. 56, 30-51)

[27] Takahasi K., Kazumasa W. (1967). *On unbiased estimates of the population mean based on the sample stratified by means of ordering,* (Ann. Inst. Statist. Math. 20, 1-31)

[28] Tukey J. (1958). *Bias and confidence in not quite large samples, abstract,* (Ann. Math. Statist. 29, 614-)

[29] Witting H., Müller-Funk U. (1995). *Mathematische Statistik II,* (B.G. Teubner Stuttgart)

[30] Zhao X., Chen Z. (2002). *On the ranked set sampling M-estimates for symmetric location families,* (Ann. Inst. Stat. Math. 54, No.3. 626-640)

Curriculum Vitae

19.09.1977	Born in Bucharest, Romania
1984-1996	German School and High School "Hermann Oberth" in Bucharest
1996-2000	Study at the Faculty of Mathematics of the University of Bucharest, in the Department of Applied Mathematics
July 2000	Diploma in mathematics at the University of Bucharest, field of the diploma thesis: mathematical statistics
2000-2004	Ph.D. student at the Institute of Mathematical Stochastics, University of Göttingen, supervisor: Prof. Dr. M. Denker
March 2001-April 2002	Assistant at the Department of Medical Statistics, University of Göttingen
April 2002-Oct. 2002	Assistant at the Institute of Mathematical Stochastics, University of Göttingen
2002-2004	Student in the Center of Statistics, University of Göttingen