

Received: December 10, 2020
Accepted: December 18, 2020

ISSN 1857–9027
e-ISSN 1857–9949
UDC: 512.53:519.243
DOI:10.20903/csnmbs.masa.2020.41.2.164

Original scientific paper

INVERSE SAMPLING DESIGNS

Žaneta Popeska

UKIM, Faculty of Computer Sciences and Engineering, Skopje, Republic of Macedonia
e-mail: zaneta.popeska@finki.ukim.mk

Dedicated to Professor Gjorgji Čupona

Using the algebraic definition of a sampling design introduced in [6], and the notion of quotient sampling designs described in [5] and [6] we present the definition of inverse sampling designs and examine some of their properties.

Key words: semigroup, free semigroup, epimorphism, sampling design

INTRODUCTION

In the sampling theory, as a part of mathematical statistics, that has been developed for several decades, one can find different approaches in selecting a sample from a population. The discrete and very often finite nature of population that is of interest in the theory of sampling design, enables use of finite algebraic structures in research in this area of statistics. In [6] we have examined the algebraic structure of sampling designs, gave a unified formal definition of the notion of sampling design that opened the way of construction of new interesting designs with some better properties in terms of their use in statistical inference. In [5] we have shown how to construct a quotient design of a given sampling design. In this paper we present the results about the opposite task, namely, we construct inverse sampling designs that can be associated to a given sampling design.

Further on, when it is clear from the contest, we will use only the word design instead of sampling design.

At the beginning we present some preliminaries. In sections two and three we state the unified definition of a sampling design as an algebraic structure and the definition of a quotient design, give some examples and state some already published re-

sults about quotient designs. In section four a construction and characterization of inverse designs is given.

Let $B = \{b_1, \dots, b_N\}$ be a finite set (called population), $S=S(B)$ be a semigroup generated by B , and $U=U(B)$ be a free semigroup generated by B . The elements of $U(B)$ will be denoted by $\sigma, \tau, \omega, \dots$ and the elements of the semigroup $S(B)$ by s, t, u, \dots

Let $\sigma \in U, \sigma = b_1 \dots b_n$, for $b_i \in B$. For a given $b \in B$, we say that $b \in \sigma$, if $b = b_i$ for some $1 \leq i \leq n$. The length $L(\sigma)$ of σ is n . We define the content $C(\sigma)$ of $\sigma \in U$, by

$$C(\sigma) = \{b \mid b \in \sigma\}.$$

If $S(B)$ is a semigroup generated by B , then there exists a unique homomorphism (which is an epimorphism) $\psi: U(B) \rightarrow S(B)$ such that $\psi(b) = b$ for each $b \in B$ ([1]). From now on we will use the symbol ψ only for this epimorphism.

SAMPLE AS AN ELEMENT OF A SEMIGROUP

In this section we give the definition of a sampling design via semigroups and give some examples that show how some known sampling designs can be represented in terms of this definition.

Let $B = \{b_1, \dots, b_N\}$ be an identifiable population and $S(B)$ a semigroup generated by B .

Definition 2.1 A *sampling design* over the population B and the semigroup S is an ordered triple $\mathbf{P} = (B, S, p)$, where $p: S(B) \rightarrow \mathbb{R}$ is a real function such that:

- i) For each $s \in S, p(s) \geq 0$; and
- ii) $\sum_{s \in S(B)} p(s) = 1$.

The semigroup $S(B)$ is called a *sampling set* and the function p - a *design function*. The elements of $S(B)$ are called S - *samples* over B , i.e., samples over B in the semigroup S .

A *carrier* of the design \mathbf{P} is the set

$$S_p = \{s | s \in S(B), p(s) > 0\}.$$

A unit $b \in B$ belongs to a sample $s \in S$, denoted by $b \in s$, if $s = a_1 \cdots a_n$ for $a_1, \dots, a_n \in B$ and there is an i such that $1 \leq i \leq n$ and $b = a_i$. In other words $b \in s$ if and only if there is a $\sigma \in U$, such that $b \in \sigma$ and $\psi(\sigma) = s$.

A sampling design $\mathbf{P} = (B, S, p)$ is called a *regular* design if for each $b \in B$, there is an $s \in S_p$ such that $b \in s$.

We say that a sampling design $\mathbf{P} = (B, S, p)$ is *finite* design if the carrier of \mathbf{P} , S_p is a finite set.

The *content* $C(s)$ of s , is defined by

$$C(s) = \{C(\sigma) | \sigma \in U, \psi(\sigma) = s\}.$$

By this definition we have that if $b \in B, s \in S$, then $b \in s$ if and only if $b \in U \{C | C \in C(s)\}$.

The *length* $L(s)$ of s , is defined by

$$L(s) = \{n | n = L(\sigma), \psi(\sigma) = s\}.$$

In other words, the length of s is the set of natural numbers that are lengths of the representations of s as a product, i.e., that are lengths of all $\sigma \in U$, such that $\psi(\sigma) = s$.

We say that a pair (B, S) satisfies the condition for *uniqueness of content* (or *length*) if and only if $C(s)$ (or $L(s)$) is a set with one element, for each $s \in S(B)$.

The following examples illustrate the representation of different sampling designs, known in literature, dealing with sampling designs, in terms of Definition 2.1.

Example 2.1. In [2] a sample is defined as a finite sequence of units of a population with replications – *ordered sampling design with replications*. It can be represented by a design $\mathbf{P} = (B, U, p)$ where U is a free semigroup generated by B , whose elements (samples) are finite ordered sequences of B with replication. By the definition, it follows that the pair (B, U) satisfies the condition for uniqueness of contents

and the condition for uniqueness of length. The same representation $\mathbf{P} = (B, U, p)$ is valid for an *ordered sampling design with replications of fixed length*, where if $p(s) > 0$, the length of s is some fixed number m .

Example 2.2. In [3] a sample is defined as a finite sequence of units of a population without replications – *ordered sampling design without replications*. This design is of the form $\mathbf{P} = (B, S, p)$, where S is the semigroup generated by B in which the following identities hold for each $x, y \in S$:

$$x^2 = x \quad \text{and} \quad xyx = xy.$$

The condition for uniqueness of contents is satisfied but not the condition for uniqueness of length, although there is a unique canonical representation of each element of S as a product of units and can be used for definition of unique content and length.

Example 2.3. In [4] a sample is defined as a subset of B – *unordered sampling design without replication*. According to our definition, this design can be represented by $(B, M(B), p)$ where $M(B)$ is a free semi-lattice generated by B , i.e. the semigroup where, for each $x, y \in B$, the following identities hold

$$x^2 = x \quad \text{and} \quad xy = yx.$$

The condition for uniqueness of content is satisfied but the condition for uniqueness of length in general is not satisfied, although as in the previous example there is a unique canonical representation for each element of $M(B)$ that can be used for definition of unique content and length.

Example 2.4. A sampling design where the sample is defined as a multi subset of B , is called, *an unordered sampling design with replications*. By Definition 2.1, a design of this type over a population B is of the form (B, N, p) where $N = N(B)$ is a free commutative semigroup generated by B , i.e. the semigroup in which the following identity holds for each $x, y \in B$:

$$xy = yx$$

Both conditions for uniqueness of content and length are satisfied.

Example 2.5. We give an example of a design that doesn't satisfy neither the condition for uniqueness of content nor the condition for uniqueness of length. Such a design is the design (B, S, p) where S is a semigroup in which the following identity

$$xyz = xuz$$

holds for each $x, y, z, u \in S$.

QUOTIENT DESIGNS

In this section we give the definition of quotient designs introduced in [5] and state some properties which are discussed and proven there.

Assume that $S = S(B)$ and $S' = S'(B')$ are semigroups generated by finite populations B and B' , and $|B| = N$, $|B'| = N'$ with $N' \leq N$.

Theorem 3.1 Let $\mathbf{P} = (B, S, p)$ be a sampling design and let $\varphi: S \rightarrow S'$ be an epimorphism such that $\varphi(B) = B'$. If $p' = p_{\varphi}: S' \rightarrow \mathbb{R}$ is defined by

$$p'(s') = \sum_{s \in \varphi^{-1}(s')} p(s) \text{ for each } s' \in S',$$

then, $\mathbf{P}' = (B', S', p')$ is a sampling design such that $S'_{p'} = \varphi(S_p)$. ■

We say that \mathbf{P}' is a *quotient design* of the design \mathbf{P} by the epimorphism φ , and denote it by \mathbf{P}_{φ} . In the same sense, we say that \mathbf{P} is an *φ -inverse design* (or just *inverse design*) of the design \mathbf{P}_{φ} .

In the above theorem and further on, for abbreviation, we use $\varphi^{-1}(s)$ instead of $\varphi^{-1}(\{s\})$.

Theorem 3.2 Any design (B, S, p') is a quotient design by some epimorphism ψ of some design (B, U, p) . ■

Proposition 3.3 Quotient design of a regular design is a regular design. ■

Proposition 3.4 Quotient design of a finite design is a finite design. ■

Proposition 3.5 For any design $\mathbf{P} = (B, S, p)$ there is a quotient design \mathbf{P}_{φ} which is regular and finite. ■

Example 3.1. Let $S' = \{1\}$ be the semigroup with one element, and $B' = S' = \{1\}$. Then there is a unique design $\mathbf{P}' = (B', S', p')$ for which $p'(1) = 1$. The design \mathbf{P}' is regular and finite and is a quotient design of any design \mathbf{P} .

Proposition 3.6 Any finite design which is not regular has a quotient design which is not regular. ■

Proposition 3.7 Any regular design which is not finite has a quotient design that is not finite. ■

INVERSE SAMPLING DESIGNS

In the previous section we gave a construction of a quotient design $\mathbf{P}' = (B', S', p')$ for a given design $\mathbf{P} = (B, S, p)$ and epimorphism $\varphi: S \rightarrow S'$, and called the design \mathbf{P} a φ -inverse design of \mathbf{P}' . In this section we will look at the opposite task, i.e., for a given design and given epimorphism, we will construct inverse designs.

Theorem 4.1 Assume that $S = S(B)$ and $S' = S'(B')$ are semigroups generated by finite populations B and B' , $|B| = N$, $|B'| = N'$ with $N' \leq N$, and $\varphi: S \rightarrow S'$ is an epimorphism such that $\varphi(B) = B'$.

Let $\mathbf{P}' = (B', S', p')$ be a sampling design and let for each $s' \in S'$,

$$p_{s'}: \varphi^{-1}(s') \rightarrow \mathbb{R}$$

be a function, such that:

a) $p_{s'}(s) \geq 0$ for each $s \in \varphi^{-1}(s')$;

b) $\sum_{s \in \varphi^{-1}(s')} p_{s'}(s) = p'(s')$.

If the function $p: S \rightarrow \mathbb{R}$ is defined by

$$p(s) = p_{\varphi(s)}(s),$$

then: $\mathbf{P} = (B, S, p)$ is a φ -inverse design of \mathbf{P}' ; $S_p \subseteq \varphi^{-1}(S'_{p'})$; and $\mathbf{P}' = \mathbf{P}_{\varphi}$.

Proof. First of all, since $\varphi(s)$ is completely determined by s , $p(s)$ is well defined real number, and it is clear that $p(s) \geq 0$. On the other hand

$$\begin{aligned} \sum_{s \in S} p(s) &= \sum_{s \in S} p_{\varphi(s)}(s) = \sum_{s' \in \varphi^{-1}(S')} \sum_{s \in \varphi^{-1}(s')} p_{s'}(s) \\ &= \sum_{s' \in S'} \sum_{s \in \varphi^{-1}(s')} p_{s'}(s) \\ &= \sum_{s' \in S'} p'(s') = 1. \end{aligned}$$

So, $\mathbf{P} = (B, S, p)$ is a design over $S = S(B)$ and

$$p'(s') = \sum_{s \in \varphi^{-1}(s')} p_{s'}(s) = \sum_{s \in \varphi^{-1}(s')} p(s).$$

This implies that \mathbf{P}' is a quotient design of \mathbf{P} by φ , and so, \mathbf{P} is φ -inverse design of \mathbf{P}' . ■

Let us note that if $s'_1, s'_2 \in S'$ and $s \in S$ are such that $s \in \varphi^{-1}(s'_1) \cap \varphi^{-1}(s'_2)$, then $s'_1 = s'_2$, meaning that $p: S \rightarrow \mathbb{R}$ is well defined. Also, if we consider the family of all functions $p_{s'}: \varphi^{-1}(s') \rightarrow \mathbb{R}$, which satisfy the conditions a) and b) in Theorem 4.1, we will obtain the family of all designs $\mathbf{P} = (B, S, p)$ that are φ -inverse of \mathbf{P}' . Particularly, if $\mathbf{P}' = \mathbf{P}_{\varphi}$, putting $p_{\varphi(s)}(s) = p(s)$, we will get the initial design \mathbf{P} .

We should emphasise that if $s' \in S'$ is such that $p'(s') = 0$, i.e., $s' \notin S'_{p'}$, then $p_{s'}(s) = 0$, for all $s \in \varphi^{-1}(s')$. Nevertheless, it is possible to have $p(s) = p_{\varphi(s)}(s) = 0$ for some $\varphi(s) \in S'_{p'}$.

The first part of the next theorem follows from the Propositions 3.5–3.7, the Example 3.1 and the Theorem 4.1.

Theorem 4.2 (i) It is possible that an inverse design of a: a) finite, b) regular, or c) finite and regular design, does not have the same property.

(ii) A design \mathbf{P}' has some of the properties a), b), or c) if and only if there is an inverse design \mathbf{P} of \mathbf{P}' that has the same property.

Proof. From the Theorem 4.1 and the Propositions 3.3 and 3.4 it follows that if there is an inverse design \mathbf{P} of \mathbf{P}' that has some of the properties a), b) or c), then the design \mathbf{P}' has the same property.

To prove the other direction of part (ii) of the theorem, for \mathbf{P}' finite or regular, we give a construction of an inverse design that is finite and inverse design that is regular and an inverse design that is finite and regular.

Let \mathbf{P}' be a finite design. We will construct a finite inverse design of \mathbf{P}' .

For all $s' \in S'_{p'}$, we choose $s_1, \dots, s_{k(s')} \in \varphi^{-1}(s')$ and real numbers $p_{1,s'}, \dots, p_{k(s'),s'} > 0$ such that

$$\sum_{i=1}^{k(s')} p_{i,s'} = p'(s').$$

Then the corresponding inverse design \mathbf{P} is finite. Let us note that if we are varying $s_1, \dots, s_{k(s')}$ and the numbers $p_{i,s'}$ over all possible values, we will get all possible finite inverse designs of \mathbf{P}' .

Let us assume now that \mathbf{P}' is a regular design. Then, there is a finite subset A' of $S'_{p'}$ such that for each $b' \in B'$, there is $s' \in A'$ such that $b' \in s'$. (We can assume that A' is the minimal set with this property, which will mean that $|A'| \leq N$, but for the following discussion this is irrelevant.) Then, for $s' \in A'$ let $a'_{1,t}, \dots, a'_{k(s'),t}$ be the elements of B' for which $a'_{t,t} \in s'$. We are looking at all units $b \in B$, for which $\varphi(b) = a'_{t,t}$ for some t . For each b with this property, we choose $s_b \in S$, such that $\varphi(s_b) = s'$ and $b \in s_b$. (This is possible since, from $a'_{t,t} \in s'$ it follows that $s' = t'a'_{t,t}q'$, where $t', q' \in S' \cup \{\lambda\}$, and λ is the empty sequence. So, the s_b we are looking for is $s_b = tbq$, for $t \in \varphi^{-1}(t')$ and $q \in \varphi^{-1}(q')$.) By $A(s')$ we denote the set of all such s_b .

Finally, we choose a function $f_{s'}: A(s') \rightarrow \mathbb{R}$ such that $p'(s') = \sum_{s \in A(s')} f_{s'}(s)$ and for each $s \in S'$, $f_{s'}(s) > 0$. Such a function $f_{s'}$ exists, since $p'(s') > 0$ and $A(s')$ are finite. For example, we can define $f_{s'}(s) = p'(s')/|A(s')|$. Then, the function $p_{s'}: \varphi^{-1}(s') \rightarrow \mathbb{R}$ is defined by

$$p_{s'}(s) = \begin{cases} f_{s'}(s) & \text{for } s \in A(s') \\ 0 & \text{otherwise} \end{cases}$$

Any inverse design \mathbf{P} obtained in this way is regular.

Note that by taking different choices for the sets $A(s')$ as well as different functions $f_{s'}$, we will get different regular inverse designs.

At the end, if \mathbf{P}' is finite and regular, we can take $A' = S'_{p'}$, so, any regular inverse design \mathbf{P} , obtained by the previous discussion, is finite too. ■

Proposition 4.3 For arbitrary design $\mathbf{P}'(B', S', p')$ there is a φ -inverse design $\mathbf{P}(B, S, p)$ of \mathbf{P}' such that the function $\tilde{\varphi}: S_p \rightarrow S'_{p'}$, induced by φ , is a bijection. The design \mathbf{P} is finite if and only if \mathbf{P}' is finite.

Proof. We obtain such a design if for each $s' \in S'_{p'}$ we choose exactly one $s \in \varphi^{-1}(s')$ and put $p_{s'}(s) = p'(s')$ and $p_{s'}(t) = 0$ for any other $t \in \varphi^{-1}(s'), t \neq s$. If \mathbf{P}' is finite, \mathbf{P} is finite too, since $|S_p| = |S'_{p'}|$. ■

With the next example we show that the last conclusion of the previous proposition does not hold for regular designs.

Example 4.1. Let $B = \{b_1, \dots, b_N\}$, $U = U(B)$, $B' = \{b_1, \dots, b_{N-1}\}$, $U' = U'(B')$ and let $\varphi: U \rightarrow U'$ be the epimorphism generated by $b_i \mapsto b_i$ for $i \leq N-1$ and $b_N \mapsto b_{N-1}$. Let $\mathbf{P}' = (B', S', p')$ be a regular design. Note that U' is a subsemigroup of U , so $U'_{p'}$ is a subset of U . If we take $U_p = U'_{p'}$, $p(s) = p'(s)$, for $s \in U_p$ and $p(s) = 0$ for $s \notin U_p$, we obtain an inverse design \mathbf{P} of the design \mathbf{P}' which is not regular even though $\tilde{\varphi}: U_p \rightarrow U'_{p'}$ is a bijection.

The validity of the next proposition is a consequence of the Theorem 4.2.

Proposition 4.4 There is a unique φ -inverse design of a given design $\mathbf{P}' = (B', S', p')$ if and only if $\varphi^{-1}(s')$ has only one element for each $s' \in S'_{p'}$. If this condition is not satisfied, then there are infinitely many φ -inverse designs of the design \mathbf{P}' . ■

Proposition 4.5 Any φ -inverse design of a finite design \mathbf{P}' is finite if and only if $\varphi^{-1}(s')$ is finite for all $s' \in S'_{p'}$.

Proof. Let $s'_0 \in S'_{p'}$ be such that $\varphi^{-1}(s'_0)$ is an infinite set and let $A = \{s_1, \dots, s_n, \dots\} \subseteq \varphi^{-1}(s'_0)$ be such that for $i \neq j, s_i \neq s_j$. We choose a sequence of positive real numbers p_1, \dots, p_n, \dots such that

$$\sum_{n=1}^{\infty} p_n = p'(s'_0).$$

If \mathbf{P} is a φ -inverse design of \mathbf{P}' such that $p(s_i) = p_i$, then \mathbf{P} is not finite since $A \subseteq S_p$. ■

For a similar characterisation of regular designs as the previous property, we need to introduce the following notion.

Let $\mathbf{P} = (B, S, p)$ be a sampling design. We say that a subset $T \subseteq S$ is a *regular subset* of S if for each $b \in B$, there is a $t \in T$, such that $b \in t$.

The subset $T \subseteq S$ is a *minimal regular subset* of S if no other proper subset of T is regular.

If T' is minimal regular subset of S' such that $T' \subseteq S'_{p'}$, and $T \subseteq \varphi^{-1}(T')$ is such that for each $s' \in T'$, $|T \cap \varphi^{-1}(s')| = 1$, then T is regular subset of S_p .

Proposition 4.6 A φ -inverse design of a regular design $\mathbf{P}' = (B', S', p')$ is regular if and only if any subset T of $\varphi^{-1}(S'_{p'})$ such that for each $s' \in S'_{p'}$, $|T \cap \varphi^{-1}(s')| = 1$, is regular in S . ■

ACKNOWLEDGMENT

My interest in algebraic structures dates back to the first year of my studies when my professor of Elementary Algebra was Professor Gjorgji Čupona. Since then, he has had a great impact on my academic and professional development.

His open mind, wide interest in different mathematical disciplines, the knowledge he unreservedly and skillfully transmitted, as well as his attitude towards his students and collaborators have been a valuable example and inspiration in my teaching and scientific work. His memorable lectures, the blackboard that at the end of the lectures looked like a carefully written part of a textbook, and mandatory consultations for all of his students at 7:30 am before the start of classes at 8:15, sparked my scientific interest in the mathematical disciplines.

When I started working on my doctoral dissertation and we discussed the problems of my interest in mathematical statistics, his suggestion was to

try to apply algebraic structures in sampling theory believing that in this way many of the questions of interest could be answered more easily.

I fill lucky and grateful that I had Professor Čupona as my teacher, academic and scientific advisor and a valuable friend.

REFERENCES

- [1] Ѓ. Чупона, Б. Трпеновски, *Предавања по алгебра*, книга II, Скопје, (1973)
- [2] C. Claes-Magnus, *Foundations of Inference in Survey Sampling*, John & Sons, Inc. (1977)
- [3] A. S. Hedajat, B.K. Sinha, *Design and inference in Finite Population Sampling*, Wiley series in probability and mathematical statistics, (1990)
- [4] Jugoslav Hajec, *Sampling From a Finite Population*, Marcel Dekker, Inc. New York, (1981)
- [5] Ź. Popeska, Quotient sampling design, *Contributions, Sec. Math. Tech. Sci.*, MASA, **XV**, 1 (1994), pp. 37–48
- [6] Ж. Попеска, *Алгебарско комбинаторни методи во теорија на примерок и експериментален дизајн*, Докторска дисертација, ПМФ, УКИМ Скопје, (1998)

ИНВЕРЗНИ ДИЗАЈНИ НА ПРИМЕРОК

Жанета Попеска

УКИМ, Факултет за информатички науки и компјутерско инженерство
Скопје, Република Македонија

Посветено на професор Ѓорѓи Чупона

Користејќи ја дефиниција на примерок во термини на алгебарски структури, воведена во [6], како и поимот за фактор план на примерок опишан во [5] и [6] дефинираме инверзен план на примерок и ги испитуваме некои од својствата кои ги имаат овие планови.

Клучни зборови: полугрупи, слободна полугрупа, епиморфизам, план на примерок.

Мојот интерес за алгебарските структури потекнува уште од првата година на моите студии по математика, кога професор по елементарна алгебра ми беше професорот Ѓорѓи Чупона. Од тогаш па се до денес, тој имаше големо влијание и беше дел од мојот научен и професионален развој. Неговата сестраност како математичар, знаењето кое безрезервно и умешно го пренесуваше, како и неговиот однос кон студентите и соработниците претставуваа пример и инспирација во мојата наставна и научна работа. Неговите незаборавни предавања, таблата која на крајот на часовите изгледаше како грижливо напишан дел од учебник и задолжителните консултации со сите студенти во вторник во 7:30, пред почетокот на часовите во 8:15 во Математичкиот амфитеатар на ПМФ, го побудија мојот научен интерес во математичките дисциплини. Кога почнав да работам на мојата докторска дисертација и ги дискутиравме проблемите од мојот интерес во математичката статистика, негова сугестија беше да се обидам да применим алгебарски структури во теоријата на примерок верувајќи дека на тој начин многу од прашањата од интерес ќе можат поедноставно да се одговорат.

Се чувствувам среќна и благодарна што бев студент и соработник на професорот Чупона, што беше мој учител, академски и научен ментор и извонреден пријател.