Sampling Space and Statistics on a Sample as an Element of a Semigroup

Ž aneta Popeska
Faculty of Natural Sciences and Mathematics ,
Skopje, Republic of Macedonia
E-mail: zaneta@pmf.ukim.edu.mk

1. Sampling design

Let $B=\{b_1, b_2,..., b_N\}$ be a finite population, $\boldsymbol{S}=\boldsymbol{S}(B)$ or $\boldsymbol{S}=<B>$, a semigroup generated by B, and $\boldsymbol{U}(B)$ a fee semigroup generated by B. The elements of $\boldsymbol{U}(B)$ will be denoted by σ, τ , ω, . . ., and the elements of a semigroup $\boldsymbol{S}(B)$ by ***s, t, u, ...*** We define **a contents of** $s \in U$, by $C(\sigma)=\{b| b\in\sigma\}$ and **a length of** s by $L(\sigma)=n$ iff $\sigma=b_1b_2....b_n$, for $b_i\in B$.
If $\boldsymbol{S}(B)$ is a semigroup generated by B, then there exists a unique homomorphism (which is epimorphism) $\psi:\boldsymbol{U}(B)\rightarrow\boldsymbol{S}(B)$ for which $\psi(b)=b$ for each $b\in B$. From now on we will use the symbol ψ only for this epimorphism. We define **a contents of $\boldsymbol{s}\in\boldsymbol{S}$**, by $C(\boldsymbol{s})=\{C(\sigma)| \psi(\sigma)=\boldsymbol{s}\}$ and **a length of $\boldsymbol{s}$** by $L(\boldsymbol{s})=\{n|n=L(\sigma),\psi(\sigma)=\boldsymbol{s}\}$.
Let $p:\boldsymbol{S}(B)\rightarrow[0,1]$ be a real function.
Given B, $\boldsymbol{S}(B)$, the triple $\mathbf{P} =(B,\boldsymbol{S}(B),p)$ is called a sampling design if

i) $\Sigma\, p(\boldsymbol{s}) = 1$

ii) $\forall b \in B, \exists\, \boldsymbol{s} \in \boldsymbol{S}(B)$, such that $b \in \boldsymbol{s}$ and $p(\boldsymbol{s})>0$.

The semigroup $\boldsymbol{S}(B)$ is called **a sampling set**, and the real function p - **a design function**. The elements of $\boldsymbol{S}$ will be called **$\boldsymbol{S}$ - samples on** B. We say that the unit $b\in B$ belongs to a sample $\boldsymbol{s}\in\boldsymbol{S}$, and write $b\in\boldsymbol{s}$ if $\boldsymbol{s}$ can be written as a product of elements of B in which b appears, i. e. $\boldsymbol{s}=b_1\cdot b_2 \cdot ...\cdot b_n\in\boldsymbol{S}(B)$, for $b_1,b_2, ...b_n \in B$ and $\exists i$, $1\le i\le n$, $b = b_i$. In other words $b\in\boldsymbol{s}$ if and only if (iff) there is $\sigma\in\boldsymbol{U}$ such that $b\in C(\sigma)$ and $\psi(\sigma)=\boldsymbol{s}$.

2. Sampling space

Let B be a finite population and $\mathbf{Y}:B\rightarrow\Re$ a mapping . The vector $\mathbf{Y}=(Y_1,Y_2, ...Y_N)$, where $Y_i =\mathbf{Y}(b_i)$ is called **a population parameter**. Let $\boldsymbol{U}$ be the free semigroup and $\boldsymbol{S}$ a semigroup generated by B and $\psi:\boldsymbol{U}\rightarrow\boldsymbol{S}$ the epimorphism defined earlier. If we consider $\sigma\in\boldsymbol{U}$ as a mapping $\sigma: b_i\rightarrow\sigma(b_i)$, let $\mathbf{y}=\mathbf{Y}\sigma=(Y_{\sigma(b1)}, ...,Y_{\sigma(bn)})$. For $\boldsymbol{s}\in\boldsymbol{S}$ we define $\mathbf{Y}\boldsymbol{s}=\{\mathbf{Y}\sigma \mid \sigma\in\psi^{-1}(\{\boldsymbol{s}\}) \}$ and $(\boldsymbol{s}:\mathbf{Y})=\{(\sigma,\mathbf{Y}\sigma) \mid \sigma\in\psi^{-1}(\{\boldsymbol{s}\})\}$. Let $\Delta(\boldsymbol{S})=\{(\boldsymbol{s},\mathbf{y})| \boldsymbol{s}\in\boldsymbol{S}_p, \mathbf{y}\in\Re^n, n\in L(\boldsymbol{s}), \exists\mathbf{Y}\in\Re^N, \mathbf{y}\in\mathbf{Y}\boldsymbol{s}\}$ and $\Delta^*(\boldsymbol{S})=\{(\boldsymbol{s}:\mathbf{Y})|\boldsymbol{s}\in\boldsymbol{S},\mathbf{Y}\in\Re^N)$ We define a relation ~ on $\Delta(\boldsymbol{U})$ by

$(\sigma,\mathbf{y})\sim(\tau,\mathbf{z})$ iff ($\psi(\sigma)=\psi(\tau)$ and $\sigma_i=\tau_j \Rightarrow y_i=z_j$).

This relation satisfies the following properties:
1° $(\sigma,\mathbf{y})\sim(\tau,\mathbf{z}) \Rightarrow(\tau,\mathbf{z})\sim (\sigma,\mathbf{y})$.♦
2° $\forall(\sigma,\mathbf{y})\in \Delta(\boldsymbol{U})$, $(\sigma,\mathbf{y})\sim (\sigma,\mathbf{y})$.♦
3° If $\psi(\sigma)=\psi(\tau)$ and $\mathbf{y}=\mathbf{Y}\sigma$, $\mathbf{z}=\mathbf{Y}\tau$ for $Y\in\Re^N$, then $(\sigma,\mathbf{y})\sim(\tau,\mathbf{z})$.♦
4° For any σ, τ for which $\psi(\sigma)=\psi(\tau)$, holds $(\sigma,\mathbf{Y}\sigma)\sim(\tau,\mathbf{Y}\tau)\ \forall\ \mathbf{Y}\in\Re^N$.♦
5° The relation ~ does not to be transitive.♦
6° If $\psi(\sigma)=\psi(\tau) \Rightarrow C(\sigma)=C(\tau)$ then the relation ~ is an equivalence relation.♦
Let ≈ be the transitive closure of ~. Then ≈ is an equivalence relation on Δ. The set $\Delta^{\tilde{}}(\boldsymbol{S})=\Delta(\boldsymbol{U})_{/\approx}$ is called **a data space** over the semigroup $\boldsymbol{S}$ and each equivalence class $(\sigma,\mathbf{y})^{\approx}$ is **a data** and will be denoted by $[\sigma,\mathbf{y}]$.
If for each $\boldsymbol{s}\in S$, $\mathbf{y}\in\Re^n$, with $n\in L(\boldsymbol{s})$, for which there is $\sigma\in\psi^{-1}(s)$, such that ker $\sigma\subseteq$ker $\mathbf{y}$ we define a set $[\boldsymbol{s},\mathbf{y}]=\{[\tau,\mathbf{y}]|\tau\in\psi^{-1}(\boldsymbol{s})\}$ then $\Delta^{\tilde{}}(\boldsymbol{S})=\bigcup_{s\in S}[s,y]$.

7° Foe each $\sigma\in\psi^{-1}(\boldsymbol{s})$, $(\boldsymbol{s},\mathbf{Y}\boldsymbol{s})\subseteq[\sigma,\mathbf{Y}\sigma]$.♦

We say that a data d=[σ,y]∈$\Delta^{\sim}(\boldsymbol{S})$ is *consistent* with the parameter **Y**∈$\Re^N$ if and only if (σ,**Y**σ)∈[σ,**y**]. Foe each **Y**∈$\Re^N$ we define $p_Y^* : \Delta^*(S) \to [0,1]$ by

$$p_Y^*(s : Zs) = \begin{cases} p(s) & if\ Zs = Ys \\ 0 & otherwise \end{cases}.$$

8° The mapping p_Y^* is well defined and induces a probability measure on the algebra of subsets of Δ^*(S) defined by

$$P_Y^*(A) = \sum_{(s:Z)\in A} p_Y^*(s:Z) = \sum_{(s:Y)\in A} p(s) . ♦$$

Now we can define a mapping p_Y:$\Delta^{\sim}(\boldsymbol{S})\to[0,1]$ with

$$p_Y([\boldsymbol{s}, y]) = P_Y^*\{(s:V) \mid (s:V) \subseteq [\boldsymbol{s}, y]\} .$$

9° $$p_Y([\boldsymbol{s}, y]) = \begin{cases} p(s) & (\boldsymbol{s}, y) \approx (\boldsymbol{s}, Y\boldsymbol{s}) \\ 0 & otherwise \end{cases} = \begin{cases} p(y(\boldsymbol{s})) & [\boldsymbol{s}, y] = [\boldsymbol{s}, Y\boldsymbol{s}] \\ 0 & otherwise \end{cases} . ♦$$

10° The mapping p_Y induces a probability measure on the algebra of subsets of $\Delta^{\sim}(\boldsymbol{S})$, P_Y, defined for each A⊆$\Delta^{\sim}(\boldsymbol{S})$ by:

$$P_Y(A) = \sum_{[\boldsymbol{s}, y]\in A} p_Y([\boldsymbol{s}, y]). ♦$$

For each **Y**∈$\Re^N$ we define subset $\Delta^{\sim}_Y \subseteq \Delta^{\sim}(S)$ with

$\Delta^{\sim}_Y$={[σ,**y**] | [σ,**y**]∈ $\Delta^{\sim}(\boldsymbol{S})$, p_Y([σ,**y**])>0}.

11° The set $\Delta^{\sim}_Y$ is at most countable set. ♦

3. Statistics

Any mapping F:$\Delta^{\sim}\to\Re^m$ is a statistics over the design **P** =(*B*,***S***(*B*),p) . Now, having in mind that the set $\Delta^{\sim}_Y$ is at most countable and so is F($\Delta^{\sim}_Y$) one can define mathematical expectation of F , E_Y(F) by

$$E_Y(F) = \sum_{w\in F(\Delta^{\sim}_Y)} w \cdot P_Y(F^{-1}(w))$$

Then

12° $$E_Y(F) = \sum_{\substack{s\in S_p \\ for\, some \\ t\in y^{-1}(s)}} F([\boldsymbol{t}, Y\boldsymbol{t}]) \cdot p(s) . ♦$$

13° If F and G are statistics on **P** =(*B*,***S***(*B*),p) , for each Y∈$\Re^N$, and α, β∈$\Re$ if the both sides exist,

$E_Y(\alpha F+\beta G)=\alpha E_Y(F) + \beta E_Y(G)$.♦

REFERENCES

Zaneta Popeska, 1998 Algebraic and combinatorial methods in sampling theory and experimental design, Ph.D. thesis, University St. Ciril & Metodij, Skopje, Macedonia

Zaneta Popeska, 1994 Quotient sampling design, Macedonian academy of sciences &arts, XV 2, 37-48

Cassel, Claes-Magnus, 1977 Fundations of inference in survey sampling",John & Sons, Inc.

Jaroslav Hajek, 1981 Sampling from a finite population, Marcel Dekker