Statistical Inferences Through Choosing Sample Set Mapping

Abstract

Our aim in this paper is to introduce a concept trying to evaluate statistical inference techniques for all statistical inference problems with a unifying method. For each inference problem, we may define a function mapping each inference technique to a subset of the sample space. Then we can define various criterions in evaluating the size of the mapping for all techniques when they are applied for one statistical inference problem. The criterions of the size for the mapping including the volume of the mapping set and probability variation of the mapping set are considered as examples in this paper. We initiate this direction of evaluation of an inference technique in terms of the size of its corresponding mapping set is interesting whereas the use of size in our three methods still needs for further investigation.

Key words: Hypothesis testing; interval estimation; point estimation; statistical inference; statistical mapping.

1. Introduction

The general statistical inference problem arises from the fact that we do not know which of a family of distributions is the true one for describing the variability of a situation in which we make an observation. From the observation made we wish to infer something about the true distribution, or equivalently about the true parameter. There are three statistical problems, point estimation, interval estimation and hypothesis testing, dealing with inferences for a parameter θ. For each statistical problem, we aim in developing a procedure "as good (in probabilistic context) as possible".

Having made great effort, statisticians developed interesting good inference procedures for these three statistical problems. In the hypothesis testing, there are two important categories of hypothesis specification, the siginificance test and the Neyman-Pearson formulation. The Neyman-Pearson

Typeset by $\mathcal{AMS}$-TEX

formulation considers a decision problem that we want to choose one from the null hypothesis H_0 and an alternative hypothesis H_1. On the other hand, the significance test considers only one hypothesis, the null hypothesis H_0. The significance test may occur that H_0 is drawn from a scientific guess and we are vague about the alternative, and cannot easily parameterize them. Another case is that the model when H_0 is true is developed by a selection process on a subset and is to be checked with new data. Then the problem for significance test is more general than the Neyman-Pearson formulation in that when H_0 is not true there are many possibilities for the true alternative. Unfortunately the existed significance tests do not share any optimal property, for examples, the uniformly minimum variance unbiased estimator in point estimation and uniformly most powerful test in Neyman-Pearson formulation, to support them. This difficulty occurs too for the confidence interval problem where Zacks (1971, Sec 10.3)) shows that desirable results for the confidence sets such as optimality in terms of size are not attainable.

The general possibility of making inference rests in the fact that it is usually the case that a given observation is much more probable under some members of the distribution family than it is under others, so that when this observation actually occurs in practice, it becomes plausible that the true distribution belongs to the former set rather than the latter. Silvey (1975) and Lindsey (1996) pointed out that much of frequentist theory appear ad hoc because it need not be model-based, not relying on the likelihood function nor any other single unifying principle, not even having the requirement that all of the information in the data must be used. This leads the fact that these existed techniques are less convincing in the interpretation of generating a subset of plausible values in the parameter space after an observation $X = x$ has been taken.

The interest of this paper is to propose a concept trying to evaluate statistical inference techniques for all statistical inference problems with a unifying method. For each inference problem, we may define a function mapping each inference technique to a subset of the sample space. Then we can define various criterions in evaluating the size of the mapping for all techniques when they are applied for one statistical inference problem. The criterions of size for the mapping including the volume of the mapping set and probability variation of the mapping set are considered as examples in this paper. We initiate this direction of evaluation of an inference technique in terms of the size of its corresponding mapping set is interesting whereas the use of size in our methods still needs for further investigation.

2. Sample Set Mapping and Statistical Inference

2.1. Sample Set Mappings

Let $X_1, X_2, ..., X_n$ be a random sample drawn from a distribution with probability density function $f(x, \theta)$ where θ is a parameter in Θ. We denote vector $X = (X_1, X_2, ..., X_n)'$ and R_x^n the sample space of the random sample X, representing the set of all possible sample realizations. We also let $\tilde{\Theta}$ be a subset of Θ of our concern and denote Δ as the class of all subsets of R_x^n. We call the following mapping

$$\tilde{\Theta} \xrightarrow{S} \Delta \tag{2.1}$$

a sample set mapping, i.e., for each $\theta \in \tilde{\Theta}, S(\theta)$ is a subset of R_x^n. A sample set mapping may be represented as the family $\{S_\theta : \theta \in \tilde{\Theta}\}$. For convenience, we also call S_θ the sample set mapping.

Why are we considering the sample set mapping? Suppose that for a specific interest and we are working on choosing one from two sample set mappings $\{S_\theta^1 : \theta \in \tilde{\Theta}\}$ and $\{S_\theta^2 : \theta \in \tilde{\Theta}\}$. This then may be done by specifying a metric to compare sizes of sample sets S_θ^1 and S_θ^2. Our interest is to treat statistical inference problems as selection of sample set mapping from a group of sample set mappings specified by the problem.

We are going to transform the class of statistical procedures for one statistical problem, point estimation, interval estimation or hypothesis testing, to a class of sample set mappings. However, the size and the form of the mapping class is determined by the problem. Let's specify several techniques for transformations needed in statistical problems.

Definition 2.1. (a) We said that a sample set mapping S is partitioned if it is disjoint in sense that $S_\theta \cap S_{\theta^*} = \phi$ for $\theta, \theta^* \in \Theta$ and exhaustive in sense that for each $x \in R_x^n$ there is a $\theta \in \Theta$ such that $x \in S_\theta$.
(b) A sample set mapping S is said to be singular if there is a set S_0 such that $\{S_\theta : \theta \in \tilde{\Theta}\} = \{S_0\}$ (a constant type mapping).
(c) We call a sample set mapping S a level $1-\alpha$ sample set mapping if it satisfies

$$P_\theta(X \in S_\theta) = 1 - \alpha \text{ for } \theta \in \tilde{\Theta}.$$

With this, we also call S_θ a level $1-\alpha$ sample set mapping.
(d) We call a sample set mapping S a size $1-\alpha$ sample set mapping if it satisfies

$$\inf_{\theta \in \tilde{\Theta}} P_\theta(X \in S_\theta) = 1 - \alpha.$$

With this, we call S_θ a size $1-\alpha$ sample set mapping.

We are going to show that each classical statistical problem may be formulated as problems for selecting sample set mapping from a subfamily of sample set mappings of (2.1) determined by conditions in Definition 2.1. For specific, the following categories explain the transformation from statistical inference problems to classes of sample set mappings:
(1) Partitioned sample set mapping corresponds with point estimation
(2) Level $1-\alpha$ sample set mapping corresponds with interval (set) estimation
(3) Singular size $1-\alpha$ sample set mapping corresponds with hypothesis testing.

2.2. Point Estimation and Partitioned Sample Set Mapping

An estimator of parameter θ represents a function of the form:

$$R_x^n \xrightarrow{\hat{\theta}} \Theta.$$

Then, an estimator may be denoted as $\hat{\theta}(X)$ where X represents the random sample.

Theorem 2.2. (a) Let $\{S_\theta : \theta \in \Theta\}$ be a partitioned sample set mapping. If we set function $h(x) = \theta$ where θ satisfies $x \in S_\theta$, then $h(X)$ is an estimator of θ.

(b) Let $\hat{\theta}(X)$ be an estimator of parameter θ. Then $\{S_\theta : \theta \in \Theta\}$ with

$$S_\theta = \{x : \hat{\theta}(x) = \theta\} \tag{2.1}$$

is a partitioned sample set mapping.

Proof. (a) Suppose that S_θ is a partitioned sample set mapping. Then, being exhaustive and disjoint for the mapping leads to the fact that for each x, there is one and only one $\theta \in \Theta$ such that $h(x) = \theta$. Then $h(X)$ is an estimator of θ since it is a mapping from R_x^n to Θ.

(b) Let $\hat{\theta}(X)$ be an estimator of θ. For each $x \in R_x^n$, we see that $x \in S_{\hat{\theta}(x)}$. This indicates that the mapping $\{S_\theta : \theta \in \Theta\}$ is exhaustive. Suppose that this mapping is not disjoint. There exists θ and θ^* in Θ with a x in R_x^n such that $x \in S_\theta$ and $x \in S_{\theta^*}$. This violates that $\hat{\theta}(X)$ is a function. This proves the theorem. □

Not every partitioned sample set mapping $\{S_\theta : \theta \in \Theta\}$ makes estimator $h(X)$ onto parameter space Θ. On the other hand, not every estimator $\hat{\theta}(X)$ makes the partitioned sample set mapping S_θ nonempty for every $\theta \in \Theta$. Let's derive partitioned sample set mappings for some examples of point estimators.

Example 1. The maximum likelihood estimator $\hat{\theta}_{mle}(X)$ sets the partition, for $\theta \in \Theta$,

$$S_\theta = \{x : L(\theta^*, x) \le L(\theta, x) \text{ for } \theta^* \in \Theta\}.$$

If $X_1, ..., X_n$ is a random sample from exponential distribution with probability density function $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, x > 0, \theta > 0$. Maximum likelihood estimator $\hat{\theta}_{mle}(X) = \bar{X}$ set the partition, for $\theta > 0$,

$$S_\theta = \{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \sum_{i=1}^{n} x_i = n\theta \}.$$

Whenever sample point $(x_1, ..., x_n)$ with $\sum_{i=1}^{n} x_i = n\theta_0$ is observed, we will let θ_0 as our estimate of the unknown parameter θ.

On the other hand, the method of moment set the partition, for $\theta \in \Theta$,

$$S_\theta = \{x : \sum_{i=1}^{n} x_i = nE_\theta X\}.$$

If $X_1, ..., X_n$ is a random sample from Bernoulli distribution with success probability p, the estimator of method of moment is $\hat{p} = \bar{X}$ that sets the partition

$$S_p = \{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \sum_{i=1}^{n} x_i = np\}.$$

S_p is non-empty only when $p \in \{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}$. Whenever S_p is non-empty, it includes elements of number $\begin{pmatrix} n \\ np \end{pmatrix}$. □

2.3. Set Estimation and Level $1-\alpha$ Sample Set Mapping

We generally say that a family, $\Theta_X = \{\Theta_x : \Theta_x \subset \Theta \text{ and } x \in R_x^n\}$ is a $100(1-\alpha)\%$ confidence set of θ if it satisfies

$$1 - \alpha = P_\theta(\{x : \theta \in \Theta_x\}), \theta \in \Theta.$$

Theorem 2.3. (a) Let $\{S_\theta : \theta \in \Theta\}$ be a level $1-\alpha$ sample set mapping. Then $C(X)$ with $C_x = \{\theta : x \in S_\theta\}$ is a $100(1-\alpha)\%$ confidence set of θ.
(b) If Θ_X is a $100(1-\alpha)\%$ confidence set of θ, then $\{S_\theta : \theta \in \Theta\}$ with

$$S_\theta = \{x : \theta \in \Theta_x\},$$

is a level $1-\alpha$ sample set mapping.
Proof. (a) Let S_θ be a level $1-\alpha$ sample set mapping. Then $P_\theta(\theta \in C(X)) = P_\theta(X \in S_\theta) = 1-\alpha$ for $\theta \in \Theta$. Thus, $C(X)$ is a $100(1-\alpha)\%$ confidence set of θ.
(b) Let Θ_X be a $100(1-\alpha)\%$ confidence set of θ. We have $P_\theta(X \in S_\theta) = P_\theta(\theta \in \Theta_X) = 1-\alpha$ for $\theta \in \Theta$. Thus, S_θ is a level $1-\alpha$ sample set mapping. □

Let $C = (t_1(X), t_2(X))$ be a $100(1-\alpha)\%$ confidence interval for θ. When $X = x$ is observed, we choose $(t_1(x), t_2(x))$ as the possible set of the true θ. By letting

$$S_\theta = \{x : \theta \in (t_1(x), t_2(x))\},$$

it satisfies $P_\theta(X \in S_\theta) = 1 - \alpha$ for $\theta \in \Theta(= \tilde{\Theta})$. In this situation, when x in S_θ is observed we put θ in the x's possible set and when x in not in S_θ we wouldn't put θ in x's possible set. That is, the confidence set when x is observed is the collection of θ such that x is in level $1 - \alpha$ sample set mapping S_θ. The interval estimation problem is the problem for selecting a level $1 - \alpha$ sample set mapping.

On the other hand, when we have level $1 - \alpha$ sample set mapping S_θ, by letting $C_x = \{\theta : x \in S_\theta\}$, the fact $P_\theta(\theta \in C_X) = 1 - \alpha$ for $\theta \in \Theta$ indicates that C_X plays the role of level $1 - \alpha$ confidence set of θ. We also have that the problem of selecting a level $1 - \alpha$ sample set mapping is a problem of selecting level $1 - \alpha$ confidence interval.

Example 2. Let $X_1, ..., X_n$ be a random sample from Bernoulli distribution with success probability p. An approximate $100(1-\alpha)\%$ confidence interval for p is $(\bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}})$. Then, the partition is

$$S_p = \{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \bar{x} - z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \le p \le \bar{x} + z_{\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\},\ p \in (0,1).$$

Whenever sample point $(x_1, ..., x_n)$ in S_p is observed, the value p is considered as a potential true parameter point. Then S_p is the set of sample points that we favor p as the true value. □

2.4. Hypothesis Testing and Singular Size $1 - \alpha$ Sample Set Mapping

In the hypothesis testing, there are two important categories of hypothesis specification, the siginificance test and the Neyman-Pearson formulation. The Neyman-Pearson formulation considers a decision problem that we want to choose one from the null hypothesis $H_0 : \theta \in \Theta_0$ and an alternative hypothesis $H_1 : \theta \in \Theta_1$ where both H_0 and H_1 could be simple or composite hypotheses. In either case, we generally concern the problem: Is a given observation consistent with the null hypothesis H_0 or it is not? For dealing with the latter framework, our concern often is by setting a significance level α and then searching for a critical region C, which is a subset of the sample space R_x^n and satisfied that the type I error probabilities, in terms of θ in Θ_0, is less than or equal to α.

Consider the hypothesis testing problem that we have a null hypothesis $H_0 : \theta \in \Theta_0$. A test with critical region C is called a level α test if $\sup_{\theta\in\Theta_0} P_\theta(X \in C) = \alpha$. Let C^c be the complement of C. Then we have

$$\inf_{\theta\in\Theta_0} P_\theta(X \in C^c) = 1 - \alpha.$$

In this setting, the transformation

$$S_\theta = \{x : x \notin C\}$$

is a size $1 - \alpha$ sample set mapping on $\tilde{\Theta} = \Theta_0$.

Theorem 2.4. (a) Let $\{S_\theta : \theta \in \Theta_0\} = \{S\}$ be a singular size $1-\alpha$ sample mapping. By letting

$$C = S^c,$$

then C is a size α critical region for the hypothesis $H_0 : \theta \in \Theta_0$.
(b) Let C be a size α critical region for the null hypothesis $H_0 : \theta \in \Theta_0$. By letting

$$S = C^c,$$

then $\{S\}$ is a singular size $1 - \alpha$ sample set mapping.
Proof. Let S be a size $1 - \alpha$ singular sample set mapping on Θ_0. Then,

$$\sup_{\theta\in\Theta_0} P_\theta(\text{Reject } H_0) = \sup_{\theta\in\Theta_0} P_\theta(X \in C) = 1 - \inf_{\theta\in\Theta_0} P_\theta(X \in S) = \alpha.$$

This proves part (a). The proof of part (b) is skipped for that it may be analogously showed. $\square$

The complement of a singular sample set mapping may serve as a critical region.

Example 3. Again, let $X_1, ..., X_n$ be a random sample from the Bernoulli distribution. Consider the hypotheses $H_0 : p = 0.5$ vs $H_1 : p > 0.5$. We classically consider to reject H_0 when $\frac{\sqrt{n}(\bar{X}-0.5)}{0.5} \geq z_{0.05}$. We then have the partition

$$S_{p=0.5} = \{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \sum_{i=1}^{n} x_i \leq n(0.5 + z_{0.05}\frac{0.5}{\sqrt{n}})\}$$

with number of elements $\sum_{j=0}^{c} \begin{pmatrix} n \\ j \end{pmatrix}$ where $c = n[0.5 + z_{0.05}\frac{0.5}{\sqrt{n}}]$ where $[.]$ is the greatest integer function. □

Dealing with general statistical hypothesis testing problem of with null hypothesis $H_0 : \theta = \theta_0$ and with no specification of any alternative hypothesis, a significance test is a method for measuring statistical evidence against H_0 that computes p-value, the probability of an extreme set determined from a sample X with observation $X = x_0$. The classical significance test, being called the Fisherian significance test, chooses a test statistic $T = t(X)$ and determines the extreme set as values x's that $t(x)'s$ are greater than or equal to $t = t(x_0)$. The p-value may be formulated as

$$p_{x_0} = P_{\theta_0}(t(X) \geq t(x_0)).$$

We have a partition

$$S_{\theta_0} = \{x : t(x) < t(x_0)\}, \tag{2.2}$$

where, in this case, $\alpha = p_{x_0}$ and $\tilde{\Theta} = \{\theta_0\}$. This includes all non-extreme samples in S_{θ_0}.

There are several properties for various θ-favorable sample sets:
(a) Consider the point estimation problem. Let $\hat{\theta}$ be an estimator with θ-favorable sample set S_θ. Then, for every $x \in R_x^n$, there exists one and only one $\theta \in \Theta$ such that x is in S_θ. With this, we have

$$S_\theta \cap S_{\theta^*} = \phi \text{ for } \theta, \theta^* \in \Theta \text{ with } \theta \neq \theta^*. \tag{2.3}$$

(b) For the confidence interval problem, the property of (2.3) is not guranteed for the induced θ-favorable sample sets.

Let S_θ be the θ-favorable sample set for a statistical technique. For every statistical inference problem that we are working on deciding a θ-favorable sample set, there may exist many choices of possible partition S_θ that all fulfill the confidence restriction.

3. Equivariant Sample Set Mapping

With the fact that the concept of θ-favorable set may be considered as a generalization of location point to a sample location set, we may expect that a location-scale equivariance property is satisfied for the θ-favorable set. In this section, let's denote θ_X as the parameter of θ with respect to random sample X.

Definition 3.1. Let's redenote the θ-favorable sample set for random variable X by $S_{\theta_x}(X)$. We say that $S_{\theta_x}(X)$ is equivariant if it satisfies $S_{\theta_{ax+b}}(aX+b) = aS_{\theta_x}(X) + b$ for $a, b \in R$.

By saying that θ is a location parameter if it satisfies $\theta(aX+b) = a\theta + b$ and θ is a scale parameter if it satisfies $\theta(aX+b) = |a|\theta(X)$. It is interesting to see in what circumstances that the θ-favorable sample sets of the statistical inference techniques are equivariant. We say that θ_x is a location parameter if it satisfies $\theta_{ax+b} = a\theta_x + b$ and it is a scale parameter if it satisfies $\theta_{ax+b} = |a|\theta_x$, for $a, b \in R$.

Theorem 3.2. (a) Let θ be a location parameter. Then, a location-scale equivariant estimator $\hat{\theta}(X)$, i.e., $\hat{\theta}(aX+b) = a\hat{\theta}(X) + b$, leads to an equivariant θ-favorable sample set.
(b) Let θ be a scale parameter. Then a scale equivariant estimator $\hat{\theta}(X)$, $\hat{\theta}(aX+b) = |a|\hat{\theta}(X)$, leads to equivariant θ-favorable sample set.
Proof. (a) By letting θ_x as the vector parameter corresponding with the sample X, the θ-favorable sample set defined by the the location-scale estimator corresponds to $aX+b$, linear function of X, is

$$S_{\theta_{ax+b}}(aX+b) = \left\{ \begin{pmatrix} ax_1 + b \\ \vdots \\ ax_n + b \end{pmatrix} : \hat{\theta}(aX+b) = \theta_{aX+b} \right\}.$$

Since $\hat{\theta}$ is a location-scale equivariant estimator. We have $\hat{\theta}(aX+b) = a\theta + b$ and θ_x is a location parameter indicating that $\theta_{ax+b} = a\theta + b$. Then the proof is finished by the followings:

$$S_{\theta_{ax+b}}(aX+b) = \left\{ a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : \hat{\theta}(x) = \theta \right\}$$
$$= aS_{\theta_x} + b.$$

(b) When θ is a scale parameter and $\hat{\theta}(X)$ is a scale equivariant estimator. Then the θ-favorable sample set of the location-scale tranformation sample $aX+b$ is

$$S_{\theta_{aX+b}}(aX+b) = \left\{ \begin{pmatrix} ax_1+b \\ \vdots \\ ax_n+b \end{pmatrix} : \hat{\theta}(aX+b) = \theta_{aX+b} \right\}$$
$$= \left\{ a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : |a|\hat{\theta}(x) = |a|\theta \right\}$$
$$= a \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \hat{\theta}(x) = \theta \right\} + b$$
$$= aS_{\theta_x}(X) + b. \quad \square$$

This theorem indicates that a location-scale equivariant estimator of a location parameter does induce an equivariant θ-favorable sample set. On the other hand, it of a scale equivariant estimator of a scale parameter is also equivariant.

Theorem 3.3. Suppose that θ is a location parameter. We also assume that a pivotal quantity $Q(X,\theta)$ satisfies the folowings:
(a) $Q(X,\theta)$ and $Q(aX+b, a\theta+b)$ have the same distribution,
(b) $Q(aX+b, a\theta+b) = \begin{cases} Q(X,\theta) & \text{if } a>0 \\ -Q(X,\theta) & \text{if } a<0 \end{cases}$.
We then have the followings:
(1) Let q satisfy $1-\alpha = P_\theta(-q \le Q(X,\theta) \le q)$. Then the favorable sample set of the $100(1-\alpha)\%$ confidence interval inverted from the event $(-q \le Q(X,\theta) \le q)$ is equivariant.

(2) Consider the null hypothesis $H_0 : \theta = \theta_0$. Then the favorable sample set of the level α test with critical region $\{x : Q(X, \theta_0) < -q \text{ or } Q(X, \theta_0) > q\}$ is location-scale equivariant.

Proof. Let $Y = aX + b$. For (1), we want to show that $S_{\theta_{ax+b}}(aX + b) = aS_{\theta_x}(X)+b$. The favorable sample set of the $100(1-\alpha)\%$ confidence interval based on random sample $Y_1, ..., Y_n$ may have the following transformation,

$$S_{\theta_{ax+b}}(aX + b) = \left\{ y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} : -q \leq Q(aX + b, a\theta + b) \leq q \right\} \tag{3.1}$$

$$\overset{(a \geq 0)}{=} \left\{ a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : -q \leq Q(X, \theta) \leq q \right\}$$

$$= a \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : -q \leq Q(X, \theta) \leq q \right\} + b$$

$$= aS_{\theta_x}(X) + b.$$

When $a < 0$, from (3.1), we have

$$S_{\theta_{ax+b}}(aX + b) = \left\{ a \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : -q \leq -Q(X, \theta) \leq q \right\}$$

$$= a \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : -q \leq Q(X, \theta) \leq q \right\} + b$$

$$= aS_{\theta_x}(X) + b.$$

For (2), it is simply the case of (1) with $\theta = \theta_0$. $\square$

Eaxmple 4. Let $X_1, X_2, ..., X_n$ be a random sample drawn from a normal distribution $N(\mu, \sigma^2)$ where σ is known constant. Traditionally we choose $100(1-\alpha)\%$ confidence interval for μ as $C(X) = (\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$. This is drived from the pivotal quantity $Q(X, \mu) = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ which satisfies conditions (a) and (b). Then the favorable sample set of the confidence interval $C(X)$ as

$$S_\mu = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} : \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \right\}$$

is equivariant. $\square$

However, when conditions (a) and (b) hold, it doesn't guarantee that the favorable sample set of the $100(1-\alpha)\%$ confidence interval inverted from the event $(-q_1 \le Q(X,\theta) \le q_2)$ with $1-\alpha = P_\theta(-q_1 \le Q(X,\theta) \le q_2)$ is equivariant. We illustrate this with an example.

Example 5. (Continue to Example 4) Let z_1 and z_2 satisfy $P(z_1 \le Z \le z_2) = 1-\alpha$ and $P(Z \le z_1) \ne P(Z \ge z_2)$. Then $C(X) = (\bar{X} + z_1\frac{\sigma}{\sqrt{n}}, \bar{X} + z_2\frac{\sigma}{\sqrt{n}})$ is a $100(1-\alpha)\%$ confidence interval for μ. The favorable sample set of $Y = aX + b$ satisfies the folowings:

$$\begin{aligned} S_{a\mu+b} &= \left\{ y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} : \bar{y} + z_1\frac{\sigma_y}{\sqrt{n}} \le \mu_y \le \bar{y} + z_2\frac{\sigma_y}{\sqrt{n}} \right\} \\ &= \left\{ a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : a\bar{x} + z_1\frac{|a|\sigma}{\sqrt{n}} \le a\mu \le a\bar{x} + z_2\frac{|a|\sigma}{\sqrt{n}} \right\} \\ &\overset{(a\le 0)}{=} \left\{ a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : a\bar{x} - z_1\frac{a\sigma}{\sqrt{n}} \le a\mu \le a\bar{x} - z_2\frac{a\sigma}{\sqrt{n}} \right\} \\ &= \left\{ a\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b : \bar{x} - z_2\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} - z_1\frac{\sigma}{\sqrt{n}} \right\} \end{aligned} \tag{3.2}$$

(3.2) is equal to $aS_\mu + b$ only if

$$P(Z \le -z_2) = P(Z \le z_1) \text{ and } P(Z \ge -z_1) = P(Z \ge z_2). \tag{3.3}$$

However, since $z_1 < 0$ and $z_2 > 0$ whenever $1-\alpha > 0$, we have $P(Z \le -z_2) = P(Z \ge z_2) \ne P(Z \le z_1)$ and $P(Z \ge -z_1) = P(Z \le z_1) \ne P(Z \ge$

z_2). This contradicts (3.3) and then the favorable sample set of $C(X)$ is not location-scale equivariant. $\square$

Theorem 3.4. Suppose that θ is a scale parameter of random variable X in sense that the parameter for $aX+b$ is $|a|\theta$. We also assume that a pivotal quantity $Q(X,\theta)$ satisfies the followings:
(a) $Q(X,\theta)$ and $Q(aX+b,a\theta+b)$ have the same distribution,
(b) $Q(aX+b,a\theta+b)=Q(X,\theta)$ for $a\neq 0$.
We then have the followings:
(1) Let q_1 and q_2 be any values satisfying $1-\alpha=P_\theta(q_1\leq Q(X,\theta)\leq q_2)$. Then the favorable sample set of the $100(1-\alpha)\%$ confidence interval inverted from the event $(q_1\leq Q(X,\theta)\leq q_2)$ is equivariant.
(2) Consider the null hypothesis $H_0:\theta=\theta_0$. Then the favorable sample set of the level α test with critical region $\{x:Q(X,\theta_0)<q_1 \text{ or } Q(X,\theta_0)>q_2\}$ is equivariant.
Proof. We only show case (1) where (2) in its special case. Let $Y=aX+b$. We want to show that $S_{\theta_{aX+b}}(aX+b)=aS_{\theta_X}(X)+b$. The favorable sample set of the $100(1-\alpha)\%$ confidence interval based on random sample $Y_1,...,Y_n$ may have the following transformation,

$$S_{\theta_{aX+b}}(aX+b)=\left\{y=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}:q_1\leq Q(aX+b,a\theta+b)\leq q_2\right\}$$

$$=\left\{a\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}+b:q_1\leq Q(X,\theta)\leq q_2\right\}$$

$$=aS_{\theta_X}(X)+b. \quad \square$$

Example 6. Let $X_1,..,X_n$ be a random sample from $N(\mu,\sigma^2)$ where μ and σ are both unknown. It is seens that $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ where $S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ is a pivotal quantity satisfying conditions (a) and (b). The interval

$$C(X)=(\frac{(n-1)S^2}{\chi_2^2},\frac{(n-1)S^2}{\chi_1^2}) \tag{2.5}$$

with $1-\alpha = P(\chi_1^2 \le \chi^2(n-1) \le \chi_2^2)$ is a $100(1-\alpha)\%$ confidence interval for σ^2. Then the σ-favorable sample set induced from this confidence interval is equivariant. $\square$

4. Volume-based Size of θ-Favorable Sample Set

For each statistical inference problem,, we face the problem of selection one from many available procedures. With sample set mapping, this problem turns to the problem of selection one from a class of avalibale sample set mappings.

The conditional distribution of the random sample $X = (X_1, ..., X_n)$ given sample space mapping S_θ is

$$f_{X|S_\theta}(x,\theta) = \frac{1}{\int_{S_\theta} f_X(x,\theta)dx} f_X(x,\theta), x \in S_\theta.$$

In case that $P_\theta(X \in S_\theta) > 0$, we let

$$f_{X|S_\theta}(x,\theta) = \frac{1}{P_\theta(X \in S_\theta)} f_X(x,\theta), x \in S_\theta.$$

What do we have for comparison of inference procedures from the structure of sample set mapping? With the above, we have an induced statistical model as:

[i] Sampling model: $X = (X_1, ..., X_n)$ is a random sample,

[ii] Induced probability model $\{f_{X|S_\theta}(x,\theta), \theta \in \Theta, x \in R_x^n\}$.

Size, in some way, for the sample set mapping is statistically appropriate for mapping selection. There are three potentially desirable techniques for computing the size of a sample set mapping:

(a) Volume of the sample set mapping S_θ provides one way for comparison,

(b) Covariance matrix of X restricted on its induced space S_θ provides another way for comparison. This approach involves the induced probability model $f_{X|S_\theta}(x,\theta)$,

(c) The third way is comparison of densities of X on S_θ that measures the probability at sample point x when θ is true. This implements the concept of Silvey(1975) and Lindsey(1996) for developing probable or plausible statistical procedures.

Without involving the density function, the size of the θ-favorable sample set may be measured with its volume for evaluation or comparison. In the attempt of proposing standard technique for measuring the size of the θ-favorable sample sets, volume is a most intuitive way to be mentioned.

Example 7. (a) Consider the example that we have a random sample drawn from Bernoulli distribution. The sample space of this random sample has elements with total number $\sum_{j=1}^{n} \begin{pmatrix} n \\ j \end{pmatrix} = 2^n$ which is finite. Hence any statistical inference procedure corresponds to a θ-favorable sample set with size, number of elements, less than or equal to 2^n. Any two statistical inference proposals may be compared through their corresponding θ-favorable sample sets in terms of element numbers.
(b) For continuous distribution, consider a simple situation that we have a random sample of size two, X_1, X_2 from a distribution with probability density function $f(x)$ and we consider a hypothesis $H_0 : f(x) = \frac{3}{2}x^2, -1 < x < 1$. The size of the density increases as $|x|$ increases from zero to one. With this fact, two seems to be reasonable tests are:
Test 1: rejecting H_0 if $|x_1 x_2| \leq 0.2057$
Test 2: rejecting H_0 if $|x_1 + x_2| \leq 1.0$.
The critical values are chosen such that both are level 0.05 tests. In this situation, the θ-favorable sample sets for Test 1 and Test 2, respectively, are

$$S_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : |x_1 x_2| \geq 0.2057 \right\} \text{ and } S_2 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : |x_1 + x_2| \geq 1.0 \right\}.$$

We then have areas of these two θ-favorable sample sets as $Area(S_1) = 1.8761$ and $Area(S_2) = 2.0$. When we set volume as a tool for comparison of θ-favorable sample sets, we would say that Test 1 is better than Test 2 in sense that it has θ-favorable sample set with smaller volume than the other one. □

Discrete and continuous disributions with sample space, for the random sample, of finite elements or finite volume, all their θ-favorable sample sets are with finite elements or finite volumes so that a comparison of these sets is achievable. How is it when the sample space for the random sample is with infinite volume? For specific, in this sample space, is a finite width confidence interval or acceptance region of a test must have θ-favorable sample set of finite volume?

Example 8. The sample space for a random sample from a normal distribution $N(\mu, \sigma_0^2)$, where σ_0 is known, is with infinite volume. The $100(1-\alpha)\%$ confidence interval of μ is $C(X) = (\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}})$ which has finite width. By letting $z^* = z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}$, the θ-favorable sample set is

$$S_\mu = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : n(\mu - z^*) \le \sum_{i=1}^{n} x_i \le n(\mu + z^*) \right\}.$$

This is a set of hyperplanes $\sum_{i=1}^{n} x_i = c, n(\mu - z^*) \le c \le n(\mu + z^*)$, obviously a set of infinite volume. Same conclusion will be drawn for the confidence interval of σ^2 in Example 6. In this situation, a comparison of the θ-favorable sample sets is generally impossible. □

Although it is not generally that we can compare θ-favorable sample sets through their volumes when the sample space of the random sample is unbounded. However, there is a situation where sample sets covered the random sample with the same coverage probability that this difficulty may be conquered.

We have defined that S_θ is a level $1-\alpha$ sample set mapping on $\tilde{\Theta}$ if it satisfies

$$1 - \alpha = P_\theta\{X \in S_\theta\} \text{ for } \theta \in \tilde{\Theta}.$$

Theorem 4.1. Suppose that two level $1-\alpha$ sample set mapping on $\tilde{\Theta}$ are $S_{\theta,a}$ and $S_{\theta,b}$ with

$$L(\theta, x_1) \ge L(\theta, x_2) \text{ for } x_1 \in S_{\theta,a} \cap S_{\theta,b}^c \text{ and } x_2 \in S_{\theta,a}^c \cap S_{\theta,b}. \tag{4.1}$$

Then $S_{\theta,a}$ has super-volume smaller than it of $S_{\theta,b}$.
Proof. Since $S_{\theta,a}$ and $S_{\theta,b}$ are both with level $1-\alpha$, we have

$$P_\theta\{X \in S_{\theta,a}\} = P_\theta\{X \in S_{\theta,b}\} \text{ for } \theta \in \tilde{\Theta}.$$

Deleting the subset common to $\{x : x \in S_{\theta,a}\}$ and $\{x : x \in S_{\theta,b}\}$ yields

$$P_\theta\{X \in S_{\theta,a} \cap S^c_{\theta,b}\} = P_\theta\{X \in S^c_{\theta,a} \cap S_{\theta,b}\}. \tag{4.2}$$

Thus, from (4.1),

$$\text{volume}(\{x : x \in S_{\theta,a} \cap S^c_{\theta,b}\} \leq \text{volume}(\{x : x \in S^c_{\theta,a} \cap S_{\theta,b}\}. \tag{4.3}$$

so, adding the volume of $\{x : x \in S_{\theta,a}\} \cap \{x : x \in S_{\theta,b}\}$ to both sides of (4.3),

$$\text{volume}(\{x : x \in S_{\theta,a}\} \leq \text{volume}(\{x : x \in S_{\theta,b}\}). \tag{4.4}$$

Then (4.4) holding for each $\theta \in \tilde{\Theta}$. $\square$

Theorem 4.1 may be applied on level $1-\alpha$ sample set mappings on $\tilde{\Theta}$ of finite or infinite volume but not at all. Here we give one example that this comparison does not work.

Example 9. Consider again that we have a hypothesis testing problem stated in Example 7. We present a case that Theorem 4.2 may not be applied. Consider the following two tests:

Test 1: rejecting H_0 if $|x_1 + x_2| \leq 1$
Test 2: rejecting H_0 if $x_1 \leq -0.9355$

(a) The acceptance region for test 1 includes the following sets:
Category 1: $S_{11} = \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : -1 < x_1 < -0.9355, 0.22 < x_2 < 1\}$, $S_{21} = \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : -0.2895 < x_1 < 0, 0.7105 < x_2 < 1\}$, $S_{11} \subset S_1$ and $S_{21} \subset S_2$.
(b) The acceptance region for test 2 includes the following sets:
Category 2: $S_{12} = \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : -0.9355 < x_1 < -0.7105, 0.0645 < x_2 < 0.22\}$, $S_{22} = \{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : -0.7105 < x_1 < -0.2895, 0.2895 < x_2 < 0.7105\}$, $S_{12} \subset S_1$ and $S_{22} \subset S_2$.

With careful checking, we may see that

$$f_{12}(x_1, x_2) > f_{12}(x_1^*, x_2^*) \text{ for } (x_1, x_2)' \in S_{11} \text{ and } (x_1^*, x_2^*)' \in S_2$$

and

$$f_{12}(x_1,x_2) < f_{12}(x_1^*,x_2^*) \text{ for } (x_1,x_2)' \in S_{12} \text{ and } (x_1^*,x_2^*)' \in S_{25}.$$

This violates the assumption in Theorem 4.1. Then this theorem is not applicable for comparison of these sample set mappings. □

We use one example to explain this point.

Theorem 4.2. Suppose that S_θ is a level $1-\alpha$ sample set mapping on $\tilde{\Theta}$ with

$$L(\theta,x_a) \geq L(\theta,x_b) \text{ for } x_a \in S_\theta \text{ and } x_b \in S_\theta^c. \tag{4.5}$$

Then S_θ has uniformly smallest volume among the class of level $1-\alpha$ sample set mappings on $\tilde{\Theta}$.

Proof. Let $S_{\theta*}$ is a level $1-\alpha$ sample set mapping on $\tilde{\Theta}$. With the discussion of (4.2), we analogously have

$$P_\theta\{X \in S_\theta \cap S_{\theta*}^c\} = P_\theta\{X \in S_\theta^c \cap S_{\theta*}\}. \tag{4.6}$$

The region on two probabilities of (4.6) cover the true parameter θ with identical probability. Now, (4.5) also indicates that $L(\theta,x_1) \geq L(\theta,x_2)$ for $x_1 \in S_\theta$ and $x_2 \in S_{\theta*}^c$, i.e., the likelihood function defined on region in left probability is greater or equal to it in the right probability. Thus,

$$\text{volume}(\{x : x \in S_\theta \cap S_{\theta*}^c\} \leq \text{volume}(\{x : x \in S_\theta^c \cap S_{\theta*}\}, \tag{4.7}$$

This further implies that

$$\text{volume}(\{x : x \in S_\theta\} \leq \text{volume}(\{x : x \in S_{\theta*}\}). \tag{4.8}$$

Then (4.8) holding for each $\theta \in \tilde{\Theta}$ and any level $1-\alpha$ sample set mapping on $\tilde{\Theta}$ leads to the theorem. □

5. Variation-based Size of Sample Set Mapping

The consideration of using volume as a standard technique for evaluation of the size of a θ-favorable sample set has a difficulty when the underlying distribution of the random sample has unbounded sample space. In this situation, the volume of an unbounded θ-favorable sample set is not finite. Although hyper-volumes may be compared through the density function in some situations of θ-favorable sample sets, however, this is not a general case to evaluate competitive θ-favorable sample sets. With this, an alternative way in measuring the size of θ-favorable sample set with the desire that it may exist with mild conditions on the underlying distribution is needed. Extending from the fact that the θ-favorable sample set may be treated as an extension of location point to location set, variation measuring the spreadness and closeness of this set may be appropriate in the representation of its size. With this consideration, the covariance matrix of the random sample restricted on the θ-favorable sample set is popularly accepted to play this role. It is interesting that this variation exists when the underlying distribution of the random sample has finite second moment.

Definition 5.1. Let S_θ be a θ-favorable sample set mapping. We call Σ_θ the variation matrix of the S_θ where

$$\Sigma_\theta = Cov_\theta[XI(X \in S_\theta)].$$

Let Y represents the n-vector having distribution of X given S_θ. Then Y has probability density function

$$f_Y(y,\theta) = \frac{1}{P_\theta(X \in S_\theta)} f_X(y,\theta), y \in S_\theta$$

whenver θ is with $P_\theta(X \in S_\theta) > 0$. In case that we consider the level $1-\alpha$ sample set mapping, then

$$f_Y(y,\theta) = \frac{1}{1-\alpha} f_X(y,\theta), y \in S_\theta.$$

The sample space of Y may varies in true parameter θ. With this transformation, we have

$$Cov_\theta(X|S_\theta) = Cov_\theta(Y).$$

Example 10. Let X_1, X_2 be a random sample of size two from a distribution with probability density function f. We consider the hypothesis $H_0 : f(x) = \frac{9}{4}x_1^2x_2^2, -1 < x_1, x_2 < 1$. The test that we reject H_0 if $|x_1 + x_2| \leq 1.0$ corresponds with the singular sample set mapping $S_0 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : |x_1 + x_2| \geq 1.0 \right\}$. The covariance matrix $\Sigma = Cov(X|S_0) = Cov(Y)$ with $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ where Y_1, Y_2 have a joint probability density function

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{9}{0.95 \times 4} y_1^2 y_2^2, (y_1, y_2) \in \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : |y_1 + y_2| \geq 1.0 \right\}.$$

With careful calculations, we see that $E(Y_1) = E(Y_2) = 0, E(Y_1^2) = E(Y_2^2) = \frac{165}{266}$ and $E(Y_1Y_2) = 0$. Then we have the population covariance matrix

$$\Sigma = \begin{pmatrix} \frac{165}{266} & 0 \\ 0 & \frac{165}{266} \end{pmatrix}.$$

On the other hand, when we consider test rejecting H_0 if $|x_1x_2| < a = 0.2057$. The corresponding singular sample set mapping $S_1 = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : |x_1x_2| \geq a \right\}$. The covariance matrix $\Sigma = Cov(X|S_1) = Cov(Y)$ with $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ where Y_1, Y_2 have a joint probability density function

$$f_{Y_1,Y_2}(y_1, y_2) = \frac{9}{0.95 \times 4} y_1^2 y_2^2, (y_1, y_2) \in \left\{ \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} : |y_1y_2| \geq a \right\}.$$

With careful calculations, we see that $E(Y_1) = E(Y_2) = 0, E(Y_1^2) = E(Y_2^2) = 0.6178$ and $E(Y_1Y_2) = 0$. Then we have the population covariance matrix

$$\Sigma = \begin{pmatrix} 0.6178 & 0 \\ 0 & 0.6178 \end{pmatrix}. \quad \square$$

Now, for two $100(1-\alpha)\%$ confidence intervals C_1 and C_2 of parameter θ, we may use the sizes of variation matrices of θ-favorable sample sets $S_{C_1}(\theta)$ and $S_{C_2}(\theta)$ as a basis to compare these two confidence intervals.

Example 11. Let $X_1, ..., X_n$ be a random sample from normal distribution $N(\mu, \sigma^2)$ where μ is unknown but σ is known. As we have noted, C_z and

C_t are the classical $100(1-\alpha)\%$ confidence intervals for μ. In this simulation, we generate data from the standard normal distribution to construct the confidence intervals and conduct replications $10,000$. In the following table, we display the simulated determinants of their corresponding sample variation matrix of the μ-favorable sample sets.

Table 1. Determinants of μ-favorable sample sets for Z and t confidence interval

n	C_z: CP	Det_z	C_t: CP	Det_t
5	0.9505	0.7616	0.9494	0.9948
10	0.9502	0.7794	0.9499	0.9989
20	0.9487	0.7790	0.9500	0.9927
30	0.9485	0.8148	0.9494	1.0280
50	0.9475	0.8242	0.9486	0.9832
100	0.9473	0.8014	0.9459	1.0236

Example 12. Let $X_1, ..., X_n$ be a random sample from the negative exponential distribution with pdf

$$f(x, \theta) = e^{-(x-\theta)}, x \geq \theta.$$

The interest is to test the hypothesis $H_0 : \theta = 0$ and we consider the following three acceptance regions

$$\text{UHLCI:}(X_{(1)} + \frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{n} - \frac{1}{2n}\chi^2_{1-\alpha}(2n), X_{(1)})$$

$$HLCI : (X_{(1)} + \frac{ln\alpha}{n}, X_{(1)})$$

$$Classical : (X_{(1)} + \frac{ln(\alpha/2)}{n}, X_{(1)} + \frac{ln(1-\alpha/2)}{n})$$

We performed a simulation study with replication $10,000$ and from the distribution under H_0. The following table display the simulated determinations of their corresponding sample variation matrices of the θ-favorable sample sets under several sample sizes.

Table 2. Determinants and associated confidences of θ-favorable sample sets for three acceptance regions

sample size	UHLCI	HLCI	Classical
$n = 5$	$2.5879e-5$ (0.9492)	$3.3627e-5$ (0.9478)	$4.4561e-5$ (0.9507)
$n = 10$	$2.2204e-14$ (0.9499)	$3.8998e-14$ (0.9468)	$4.5973e-14$ (0.9510)
$n = 20$	$3.3974e-38$ (0.9499)	$1.0160e-37$ (0.9511)	$1.1332e-37$ (0.9513)
$n = 30$	$2.6612e-66$ (0.9524)	$6.7416e-66$ (0.9478)	$9.5812e-66$ (0.9512)

References

Lindsey, J. K. (1996). *Parametric Statistical Inference.* Oxford Science Publication: New York.

Silvey, S. D. (1975). *Statistical Inference.* Chapman and Hall: London.

Zacks, S. (1971). *The Theory of Statistical Inference.* Wiley: New York.