Equivariance and Invariance

Statistical decisions or actions based on data should not be affected by simple transformations on the data or by reordering of the data, so long as these changes on the data are reflected in the statement of the decision; that is, the actions should be *invariant*.

If the action is a yes-no decision, such as in hypothesis testing, it should be completely invariant.

If a decision is a point estimate, its value is not unaffected, but it should be *equivariant*, in the sense that it reflects the transformations in a meaningful way.

We can formalize this principle by defining appropriate classes of transformations, and then specifying rules that statistical decision functions must satisfy.

The Principle of Equivariance under Transformations

We identify “reasonable” classes of transformations on the sample space and the corresponding transformations on other components of the statistical decision problem.

We will limit consideration to transformations that are one-to-one and onto.

Such transformations can easily be identified as members of a group.

Transformations

We will consider only parametric distributions $P_{X|\theta}$ for $\theta \in \Theta$.

We are interested in what happens under a transformation of the random variable $g(X)$.

We seek a transformation of the parameter $\bar{g}(\theta)$ such that $P_{g(X)|\bar{g}(\theta)}$ is a member of the same distributional family, and will the same optimal methods of inference for $P_{X|\theta}$ remain optimal for $P_{g(X)|\bar{g}(\theta)}$.

We want to identify optimal methods of inference for $P_{X|\theta}$ that will remain optimal for $P_{g(X)|\bar{g}(\theta)}$.

Transformation Groups

Definition. A *group* $\mathcal{G}$ is a nonempty set G together with a binary operation $\circ$ such that

- $g_1, g_2 \in G \Rightarrow g_1 \circ g_2 \in G$ (closure);
- $\exists\, e \in G \ni \forall\, g \in G$, $e \circ g = g$ (identity);
- $\forall\, g \in G \,\exists\, g^{-1} \in G \ni g^{-1} \circ g = e$ (inverse);
- $g_1, g_2, g_3 \in G \Rightarrow g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$ (associativity).

Notice that the binary operation need not be commutative, but from these defining properties, we can easily see that $g \circ e = e \circ g$ and $g \circ g^{-1} = g^{-1} \circ g$.

Groups

A group $\mathcal{G}$ is a structure of the form $(G, \circ)$.

Sometimes the same symbol that is used to refer to the set is used to refer to the group. The expression $g \in \mathcal{G}$ is interpreted to mean $g \in G$.

Any subset of the set on which the group is defined that is closed and contains the identity and all inverses forms a group with the same operation as the original group.

This subset together with the operation is called a *subgroup*.

We use the standard terminology of set operations for operations on groups.

Groups

A set G_1 together with an operation $\circ$ defined on G_1 *generates* a group $\mathcal{G}$ that is the smallest group $(G, \circ)$ such that $G_1 \subset G$.

If $\mathcal{G}_1$ and $\mathcal{G}_2$ are groups over G_1 and G_2 with a common operation $\circ$, the group generated by G_1 and G_2 is $(G, \circ)$, where G is the smallest set containing G_1 and G_2 so that $(G, \circ)$ is a group.

Notice that the G may contain elements that are in neither G_1 nor G_2.

Transformation Groups

If the elements of a set are transformations, function composition is a binary operation.

A set of one-to-one and onto functions with common domain together with the operation of function composition is a group, referred to as a transformation group.

A transformation group has an associated set that is the common domain of the transformations. This is called the domain of the transformation group.

Both function composition and a function-argument pair are often indicated by juxtaposition with no symbol for the operator or operation.

Transformation Groups

The expression $g^*Tg^{-1}X$ means function composition on the argument X.

We can write

$$\begin{aligned} g^*(T(g^{-1}(\tilde{X}))) &= g^*(T(X)) \\ &= g^*(a) \\ &= \tilde{a}. \end{aligned}$$

Location-Scale Families

If the conditional distribution of $X - \theta$, given $\theta = \theta_0$, is the same for all $\theta_0 \in \Theta$, then θ is called a *location parameter*.

The family of distributions $P_{X|\theta}$ is called a *location family*.

If $\Theta \subset \mathbb{R}_+$, and if the conditional distribution of X/θ, given $\theta = \theta_0$, is the same for all $\theta_0 \in \Theta$, then θ is called a *scale parameter*.

The family of distributions $P_{X|\theta}$ is called a *scale family*.

We can write the PDF of a member of a location-scale family as

$$\frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$$

Other Group Families

More generally, given a distribution with parameter θ, that distribution together with a *group* of transformations on θ forms a "group family" of distributions.

Invariant and Equivariant Functions

A function f is said to be *invariant* under the transformation group $\mathcal{G}$ with domain $\mathcal{X}$ if for all $x \in \mathcal{X}$ and $g \in \mathcal{G}$,

$$f(g(x)) = f(x).$$

We also use the phrases "invariant over ..." and "invariant with respect to ..." to denote this kind of invariance.

A function f is said to be *equivariant* under the transformation group $\mathcal{G}$ with domain $\mathcal{X}$ if for all $x \in \mathcal{X}$ and $g \in \mathcal{G}$,

$$f(g(x)) = |\partial g/\partial x| f(x).$$

Transformation Groups

A transformation group $\mathcal{G}$ defines an equivalence relation (identity, symmetry, and transitivity) for elements in its domain, $\mathcal{X}$.

If $x_1, x_2 \in \mathcal{X}$ and there exists a g in $\mathcal{G}$ such that $g(x_1) = x_2$, then we say x_1 and x_2 are equivalent under $\mathcal{G}$, and we write

$$x_1 \equiv x_2 \bmod G.$$

Sets of equivalent points are called *orbits* of $\mathcal{G}$.

It is clear that a function that is *invariant* under the transformation group $\mathcal{G}$ must be constant over the orbits of G.

A transformation group $\mathcal{G}$ is said to be *transitive* over the set $\mathcal{X}$ if for any $x_1, x_2 \in \mathcal{X}$, there exists a g in $\mathcal{G}$ such that $g(x_1) = x_2$.

Maximal Invariance

An invariant function M over $\mathcal{G}$ is called *maximal invariant* over $\mathcal{G}$ if

$$M(x_1) = M(x_2) \quad \Rightarrow \quad \exists\, g \in \mathcal{G} \ni g(x_1) = x_2.$$

Maximal invariance can be used to characterize invariance.

If M is maximal invariant under $\mathcal{G}$, then the function f is invariant under $\mathcal{G}$ if and only if it depends on x only through M; that is, if and only if there exists a function h such that for all x, $f(x) = h(M(x))$.

Any invariant function with respect to a transitive group is maximal invariant.

Maximal Invariance

The underlying concept of maximal invariance is similar to the concept of sufficiency.

A sufficient statistic may reduce the sample space; a maximal invariant statistic may reduce the parameter space.

Maximal invariant statistics have some technical issues regarding measurability, however; X being measurable does not guarantee $M(X)$ is measurable under the same measure.

Equivariant Functions

A function f is said to be *equivariant* under the transformation group $\mathcal{G}$ with domain $\mathcal{X}$ if for all $x \in \mathcal{X}$ and $g \in \mathcal{G}$,

$$f(g(x)) = g(f(x)).$$

We also use the phrases "equivariant over ..." and "equivariant with respect to ..." to denote this kind of equivariance.

Invariant and Equivariant Statistical Procedures

We will denote a general statistical decision function by T.

This is a function from the sample space (actually from the range $\mathcal{X}$ of the random variable X) to the decision space $\mathcal{A}$ (which can be considered a subset of the reals); that is, $T : \mathcal{X} \mapsto \mathcal{A} \subseteq \mathbb{R}$.

We write

$$T(X) = a.$$

We are interested in the invariance or equivariance of $T(x)$ in the context of certain transformations.

The context in which this has meaning is somewhat limited.

It has meaning in group families when the loss function is of an appropriate type.

Invariant and Equivariant Statistical Procedures

An estimator that changes appropriately (in ways that we will specify more precisely below) so that the risk is invariant under changes in the random variable is said to be equivariant.

In testing statistical hypotheses, we often denote the statistical decision function by ϕ, and define the decision space as $[0, 1]$.

We interpret $\phi(x)$ as the probability of rejecting the hypothesis for a given $x \in \mathcal{X}$.

A test that does not change under changes in the random variable is said to be invariant.

Invariant and Equivariant Statistical Procedures

We must emphasize that invariance or equivariance has meaning only in special contexts; both the family of distributions and the form of the loss function must have properties that are similar in certain ways.

The invariance or equivariance of interest is with respect to a given class of transformations.

The most common class of interest is the group of linear transformations of the form $\tilde{x} = Ax + c$.

A family of distributions whose probability measures accommodate a group of transformations in a natural way is called a *group family*.

Location-Scale Families

The group families of interest have a certain invariance with respect to a group of linear transformations on the random variable.

Formally, let P be a a probability measure on $(\mathbb{R}^k, \mathcal{B}^k)$, let $\mathcal{V} \subset \mathbb{R}^k$, and let $\mathcal{M}_k$ be a collection of $k \times k$ symmetric positive definite matrices.

The family

$$\{P_{(\mu,\Sigma)} \; : \; P_{(\mu,\Sigma)}(B) = P(\Sigma^{1/2}(B-\mu)), \text{ for } \mu \in \mathcal{V},\ \Sigma \in \mathcal{M}_k,\ B \in \mathcal{B}^k\}$$

is called a *location-scale* family.

Some standard parametric families that are group families: normal, double exponential, exponential and uniform (even with parametric ranges), and Cauchy.

Location-Scale Families

A location-scale family of distributions can be defined in terms of a given distribution on $(\mathbb{R}^k, \mathcal{B}^k)$ as all distributions for which the probability measure is invariant under linear transformations.

Whatever parameter θ may characterize the distribution, we often focus on just μ and Σ, as above, or in the univariate case, μ and σ.

In most other cases our object of interest has been a transformation on the parameter space, $g(\theta)$.

In the following, we will often denote a basic transformation of the *probability space* as $g(\cdot)$, and we may denote a corresponding transformation on the parameter space, as $\bar{g}(\cdot)$.

Transformations on the Sample Space, the Parameter Space, and the Decision Space

To study invariance of statistical procedures we will now identify three groups of transformations $\mathcal{G}$, $\overline{\mathcal{G}}$, and $\mathcal{G}^*$, and the relationships among the groups.

This notation is widely used in mathematical statistics, maybe with some slight modifications.

- Let $\mathcal{G}$ be a group of transformations that map the probability space onto itself. We write

$$g(X) = \tilde{X}.$$

Note that X and $\tilde{X}$ are random variables, so the domain and the range of the mapping are subsets of *probability spaces*; the random variables are based on the same underlying measure, so the probability spaces are the same; the transformation is a member of a transformation group, so the domain and the range are equal and the transformations are one-to-one.

$$g : \mathcal{X} \mapsto \mathcal{X}, \quad 1:1 \text{ and onto}$$

- For given $g \in \mathcal{G}$ above, let $\bar{g}$ be a 1:1 function that maps the parameter space onto itself, $\bar{g} : \Theta \mapsto \Theta$, in such a way that for any set A,

$$\Pr_{\theta}(g(X) \in A) = \Pr_{\bar{g}(\theta)}(X \in A).$$

If this is the case we say g *preserves* Θ. Any two functions that preserve the parameter space form a group of functions that preserve the parameter space. The set of all such $\bar{g}$ together with the induced structure is a group, $\bar{\mathcal{G}}$. We write

$$\bar{g}(\theta) = \tilde{\theta}.$$

$$\bar{g} : \Theta \mapsto \Theta, \quad 1:1 \text{ and onto}$$

We may refer to $\bar{\mathcal{G}}$ as the *induced* group under $\mathcal{G}$.

- For each $g \in \mathcal{G}$ above, there is a 1:1 function g^* that maps the decision space onto itself, $g^* : \mathcal{A} \mapsto \mathcal{A}$. The set of all such g^* together with the induced structure is a group, $\mathcal{G}^*$. We write

$$g^*(a) = \tilde{a}.$$

$$g^* : \mathcal{A} \mapsto \mathcal{A}, \quad 1:1 \text{ and onto.}$$

The relationship between $\mathcal{G}$ and $\mathcal{G}^*$ is a homomorphism; that is, for $g \in \mathcal{G}$ and $g^* \in \mathcal{G}^*$, if $g^* = k(g)$, then $k(g_1 \circ g_2) = k(g_1) \circ k(g_2)$.

We are interested in a probability space, $(\Omega, \mathcal{F}, \mathcal{P}_\Theta)$, that is invariant to a class of transformations $\mathcal{G}$; that is, one in which $\mathcal{P}_\Theta$ is a group family with respect to $\mathcal{G}$.

The induced groups $\bar{\mathcal{G}}$ and $\mathcal{G}^*$ determine the transformations to be applied to the parameter space and the action space.

Invariance of the Loss Function

In most statistical decision problems, we assume a *symmetry* or *invariance* or *equivariance* of the problem before application of any of these transformations, and the problem that results from applying all of the transformations.

For given classes of transformations, we consider loss functions that are invariant to those transformations; that is, we require that the loss function have the property

$$\begin{aligned} L(\tilde{\theta}, \tilde{a}) &= L(\bar{g}(\theta), g^*(a)) \\ &= L(\theta, a). \end{aligned}$$

This means that a good statistical procedure, T, for the original problem is good for the transformed problem.

Note that this is an *assumption* about the class of meaningful loss functions for this kind of statistical problem.

From this assumption about the loss function, we have the risk property

$$\mathsf{E}_\theta(g(X)) = \mathsf{E}_{\tilde{g}(\theta)}(X).$$

We have seen cases in which, for a univariate function of the parameter, the loss function is a function only of $a - g(\theta)$ or of $a/g(\theta)$; that is, we may have $L(\theta, a) = L_{\mathsf{l}}(a - g(\theta))$, or $L(\theta, a) = L_{\mathsf{s}}(a/g(\theta))$.

In order to develop equivariant procedures for a general location-scale family $P_{(\mu,\Sigma)}$ we need a loss function of the form

$$L((\mu, \Sigma), a) = L_{\mathsf{ls}}(\Sigma^{1/2}(a - \mu)).$$

Invariance of Statistical Procedures

The basic idea underlying invariance of statistical procedures naturally is invariance of the risk under the given transformations.

We seek a statistical procedure $T(x)$ that is an invariant function under the transformations.

Because if there is a maximal invariant function M all invariant functions are dependent on M, our search for optimal invariant procedures can use M.

A probability model may be defined in different ways.

There may be an equivalence between two different models that is essentially a result of a reparametrization: $\tilde{\theta} = \bar{g}(\theta)$.

A random variable in the one model may be a function of the random variable in the other model: $\widetilde{X} = g(X)$.

There are two ways of thinking of estimation under a reparametrization, both in the context of an estimator $T(X)$ of $h(\theta)$, and with the transformations defined above:

- functional, $g^*(T(X))$ estimates $g^*(h(\theta))$;

- formal, $T(g(X))$ estimates $g^*(h(\theta))$.

Functional equivariance is trivial.

This is the equivariance we expect under a simple change of units, for example.

Functional Equivariance

If X is a random variable that models physical temperatures in some application, it should not make any real difference whether the temperatures are always measured in degrees Celsius or degrees Fahrenheit.

The random variable itself does not include units, of course (it is a real number).

If the measurements are made in degrees Celsius at a time when X is the random variable used to model the distribution of the data and the estimator $T(X)$ and the estimand $h(\theta)$ relates to X in a linear fashion (if $h(\theta)$ is the mean of X, for example), and later in a similar application the measurements are made in degrees Fahrenheit, applying $g^*(t) = 9t/5+32$ to both $T(X)$ and $h(\theta)$ preserves the interpretation of the model.

Formal Equivariance

Formal equivariance, however, is not meaningful unless the problem itself has fundamentally symmetric properties; the family of probability distributions is closed under some group of transformations on the sample space one on the parameter space.

In this case, we need a corresponding transformation on the decision space.

The statistical procedure is equivariant it the functional equivariance is the same as the formal equivariance; that is,

$$T(g(X)) = g^*(T(X)).$$

Optimality

Equivariance can be combined with other properties such as minimum risk or most powerfulness.

As we have seen, there are situations where we cannot obtain these properties uniformly.

By restricting attention to procedures with properties such as equivariance or unbiasedness, we may be able to achieve uniformly best procedures.

With unbiasedness, we seek UMVU estimators and UMPU tests.

Within a collection of equivariant estimators, we would choose the one with some optimality property such as minimum risk.

The simplest and most interesting transformations are translations and scalings, and the combinations of these two, that is linear transformations.

Consequently, the two most common types of invariant inference problems are those that are location invariant (or equivariant) and those that are scale invariant (or equivariant).

A location invariant procedure is not invariant to scale transformations, but a scale invariant procedure is invariant to location transformations.

Equivariant Confidence Sets

The connection we have seen between a $1-\alpha$ confidence region $S(x)$, and the acceptance region of a α-level test, $A(\theta)$, that is

$$S(x) \ni \theta \quad \Leftrightarrow \quad x \in A(\theta),$$

can often be used to relate UMP invariant tests to best equivariant confidence sets.

Equivariance for confidence sets is defined similarly to equivariance in other settings.

For the group of transformations G and the induced transformation groups G^* and $\overline{G}$, a confidence set $S(x)$ is *equivariant* if for all $x \in \mathcal{X}$ and $g \in G$,

$$g^*(S(x)) = S(g(x)).$$

The uniformly most powerful property of the test corresponds to uniformly minimizing the probability that the confidence set contains incorrect values, and the invariance corresponds to equivariance.

An equivariant set that is $\widetilde{\Theta}$-uniformly more accurate ("more" is defined similarly to "most") than any other equivariant set is said to be a *uniformly most accurate equivariant* (UMAE) set.

There are situations in which there do not exist confidence sets that have uniformly minimum probability of including incorrect values.

In such cases, we may retain the requirement for equivariance, but impose some other criterion, such as expected smallest size (w.r.t. Lebesgue measure) of the confidence interval.