Lecture Notes for STAT-F404, author: Marc Hallin[1]

8 Inference and group invariance/equivariance

8.1 Group invariance

Let $(\mathcal{X},\ \mathcal{A},\ \mathcal{P})$ be a statistical model. Denote as $\mathbf{X}$ the observation and by P, Q, ... (or $\mathrm{P}^{\mathbf{X}}$, $\mathrm{Q}^{\mathbf{X}}, \ldots$) the distributions in $\mathcal{P}$. Throughout, we assume *identifiability*: distinct elements P, Q of $\mathcal{P}$ denote distinct distributions on $(\mathcal{X},\ \mathcal{A})$. In the parametric case $\mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$, identifiability is equivalent to $\boldsymbol{\theta} \mapsto \mathrm{P}_{\boldsymbol{\theta}}$ being injective: $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ implies $\mathrm{P}_{\boldsymbol{\theta}_1} \neq \mathrm{P}_{\boldsymbol{\theta}_2}$.

Let $\mathcal{G},\circ$ be a group of measurable transformations $g : \mathbf{x} \mapsto g(\mathbf{x})$ of $(\mathcal{X}, \mathcal{A})$.

Definition 1. *The model $(\mathcal{X},\ \mathcal{A},\ \mathcal{P})$ (or the family $\mathcal{P}$) is* invariant *under $\mathcal{G},\circ$ if, for all $\mathrm{P} \in \mathcal{P}$ and all $g \in \mathcal{G}$, there exists $\mathrm{Q} \in \mathcal{P}$ such that $\mathrm{P}^{g(\mathbf{X})} = \mathrm{Q}$.*

In view of the identifiability assumption, such Q automatically is unique. It only depends on P and g: denote it as $Q =: \bar{g}(P)$. The characteristic relation $\mathrm{P}^{g(\mathbf{X})} = \mathrm{Q}$ then takes the form

$$\mathrm{P}^{g(\mathbf{X})} = \bar{g}(\mathrm{P}^{\mathbf{X}}). \tag{8.1}$$

For any $g \in \mathcal{G}$, this defines a transformation $\bar{g} : \mathrm{P} \mapsto \bar{g}(P)$ of $\mathcal{P}$. Letting $\bar{\mathcal{G}} := \{\bar{g} : g \in \mathcal{G}\}$, it is easy to see that $\bar{\mathcal{G}},\circ$ is a *group* of transformations acting on $\mathcal{P}$—call it the *induced (by $\mathcal{G},\circ$) group.* In the parametric case ($\mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$), it is convenient to define the induced group $\bar{\mathcal{G}},\circ$ as a group acting on Θ, which leads to the following definition.

Definition 2. *The parametric model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, with $\mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$, is* invariant *under $\mathcal{G},\circ$ if, for all $g \in \mathcal{G}$ and all $\boldsymbol{\theta} \in \Theta$, there exists a (unique in view of the identifiability assumption) parameter value $\boldsymbol{\eta}$, denoted as $\bar{g}(\boldsymbol{\theta})$, such that*

$$\mathrm{P}_{\boldsymbol{\theta}}^{g(\mathbf{X})} = \mathrm{P}_{\bar{g}(\boldsymbol{\theta})}^{\mathbf{X}}. \tag{8.2}$$

Here again, letting $\bar{\mathcal{G}} := \{\bar{g} : g \in \mathcal{G}\}$, it is easy to see that $\bar{\mathcal{G}},\circ$ is a group of transformations acting on Θ—still called the induced (by $\mathcal{G},\circ$) group.

[1] With slight modifications by Davy Paindaveine and Thomas Verdebout.

Example 1: Location families.
Example 2: Location-scale families.
Example 3: Nonparametric white noise.

8.2 Generating groups

For any $\mathbf{x} \in \mathcal{X}$, let $\mathcal{G}_{\mathbf{x}} = \{g(\mathbf{x}) : g \in \mathcal{G}\}$ denote the *orbit* of $\mathbf{x}$. Assume that, for any $\mathbf{x}$, $\mathcal{G}_{\mathbf{x}} \in \mathcal{A}$. Then, these orbits constitute a measurable partition of $\mathcal{X}$. Denote by $\mathcal{A}_{\mathcal{G}}$ the sub-sigma-field of $\mathcal{A}$ generated by these orbits, and call it the *σ-field of orbits*.

Similarly, $\bar{\mathcal{G}}$ induces a partition of $\mathcal{P}$ into orbits of the form $\bar{\mathcal{G}}_{\mathrm{P}} = \{\bar{g}(\mathrm{P}) : \bar{g} \in \bar{\mathcal{G}}\}$. If this partition consists of one single orbit (for any $\mathrm{P} \in \mathcal{P}$, $\bar{\mathcal{G}}_{\mathrm{P}} = \mathcal{P}$), we say that $\mathcal{G},\circ$ is a *generating group* for the model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ (or for the family $\mathcal{P}$). In such case, indeed,

$$\text{for all } \mathrm{P}_1, \mathrm{P}_2 \in \mathcal{P}, \quad \text{there exists } g \in \mathcal{G} \text{ such that } \mathrm{P}_2^{\mathbf{X}} = \mathrm{P}_1^{g(\mathbf{X})},$$

or, equivalently,

$$\text{for all } \mathrm{P}_1, \mathrm{P}_2 \in \mathcal{P}, \quad \text{there exists } \bar{g} \in \bar{\mathcal{G}} \text{ such that } \mathrm{P}_2 = \bar{g}\mathrm{P}_1.$$

Examples 1–3: The groups of translations, of affine transformations, and of continuous strictly increasing transformations are generating groups in Examples 1–3, respectively.

8.3 Invariant - Maximal invariant

Definition 3. *An $\mathcal{A}$-measurable random variable $\mathbf{S}(\mathbf{X})$ is called an* invariant *(of $\mathcal{G},\circ$) if it is constant along the orbits of $\mathcal{G}$, that is, if*

$$\mathbf{x}' = g(\mathbf{x}) \textit{ for some } g \in \mathcal{G} \Longrightarrow \mathbf{S}(\mathbf{x}') = \mathbf{S}(\mathbf{x}).$$

If follows that $\mathbf{S}$ is an invariant iff it is $\mathcal{A}_{\mathcal{G}}$-measurable, that is, iff $\mathcal{A}_{\mathbf{S}} \subseteq \mathcal{A}_{\mathcal{G}}$.

Definition 4. *An $\mathcal{A}$-measurable random variable $\mathbf{S}^0(\mathbf{X})$ is called a* maximal invariant *(of $\mathcal{G},\circ$) if*

$$\mathbf{x}' = g(\mathbf{x}) \textit{ for some } g \in \mathcal{G} \Longleftrightarrow \mathbf{S}^0(\mathbf{x}') = \mathbf{S}^0(\mathbf{x}).$$

It follows that $\mathbf{S}^0$ is maximal invariant iff $\mathcal{A}_{\mathbf{S}^0} = \mathcal{A}_{\mathcal{G}}$—namely, iff it generates the σ-field of the orbits; hence, $\mathbf{S}$ is invariant if and only if it is $\mathbf{S}^0$-measurable.

The following properties of invariants (parametric notation is used for simplicity) are important in the framework of hypothesis testing.

Proposition 1. *The probability distribution* $\mathrm{P}^{\mathbf{S}}_{\boldsymbol{\theta}}$ *of an invariant* $\mathbf{S}$ *is constant along the orbits of* $\overline{\mathcal{G}},\circ$*: for all* $\boldsymbol{\theta} \in \Theta$ *and all* $\bar{g} \in \bar{\mathcal{G}}$*,* $\mathrm{P}^{\mathbf{S}}_{\boldsymbol{\theta}} = \mathrm{P}^{\mathbf{S}}_{\bar{g}(\boldsymbol{\theta})}$*.*

Proof. Let $\mathbf{S}$ take values in $(\mathcal{S}, \mathcal{B}_{\mathcal{S}})$. Then, for all $B \in \mathcal{B}_{\mathcal{S}}$, $\bar{g} \in \bar{\mathcal{G}}$ and $\boldsymbol{\theta} \in \Theta$,

$$\mathrm{P}^{\mathbf{S}}_{\bar{g}(\boldsymbol{\theta})}[B] := \mathrm{P}^{\mathbf{X}}_{\bar{g}(\boldsymbol{\theta})}[\mathbf{S}^{-1}(B)] = \mathrm{P}^{g(\mathbf{X})}_{\boldsymbol{\theta}}[\mathbf{S}^{-1}(B)] = \mathrm{P}^{\mathbf{S}(g(\mathbf{X}))}_{\boldsymbol{\theta}}[B].$$

But, since $\mathbf{S}$ is invariant, $\mathrm{P}^{\mathbf{S}(g(\mathbf{X}))}_{\boldsymbol{\theta}} = \mathrm{P}^{\mathbf{S}}_{\boldsymbol{\theta}}$, which completes the proof. $\square$

In particular, we have the following result.

Corollary 1. *If* $\mathcal{G},\circ$ *is a generating group, then invariants are distribution-free.*

8.4 Invariant tests

Assume that, in the statistical model $(\mathcal{X}, \mathcal{A}, \mathcal{P} = \{\mathrm{P}\})$, one wants to test $\mathcal{H}_0$ against $\mathcal{H}_1$. If this model is invariant under a group of transformations $\mathcal{G},\circ$ and if the null submodel $(\mathcal{X}, \mathcal{A}, \mathcal{H}_0)$ also is, then the testing problem is said to be invariant under $\mathcal{G},\circ$. In this framework, the *invariance principle* leads to restricting to *invariant tests* ϕ, that is, to tests satisfying $\phi(g(x)) = \phi(x)$ for any $x \in \mathcal{X}$ and any $g \in \mathcal{G}$. The results from the previous sections entail that a test ϕ is invariant if and only if ϕ is measurable with respect to a maximal invariant, $\mathbf{T} = \mathbf{T}(\mathbf{x})$, of $\mathcal{G},\circ$, i.e., if and only if $\phi(\mathbf{x}) = \varphi(\mathbf{T}(\mathbf{x}))$ for some measurable function φ. In other words, invariant tests are to be defined in the induced model $(\mathcal{T}, \mathcal{B}_{\mathcal{T}}, \mathcal{P}^{\mathbf{T}} = \{\mathrm{P}^{\mathbf{T}}\})$. Tests that are UMP in this induced model will then be said to be UMPI (*Uniformly Most Powerful Invariant*) in the original model, tests that are UMP in this induced model will then be said to be UMPUI (*Uniformly Most Powerful Unbiased Invariant*) in the original model, etc.

Let us illustrate this with two examples. Consider first the problem of testing

$$\mathcal{H}_0 : \mathbf{X} \sim f_\theta(\mathbf{x}) = f_0(x_1 - \theta, \ldots, x_n - \theta) \qquad \text{for some unspecified } \theta \in \mathbb{R}$$

against

$$\mathcal{H}_1 : \mathbf{X} \sim g_\theta(\mathbf{x}) = g_0(x_1 - \theta, \ldots, x_n - \theta) \qquad \text{for some unspecified } \theta \in \mathbb{R}$$

at level α, where f_0 and g_0 are two given (different) densities over $\mathbb{R}^n$. The null hypothesis and alternative hypothesis are thus two different location families, and the location parameter θ is a nuisance parameter. The testing problem is then invariant under the group of translations $(x_1, \ldots, x_n) \mapsto (x_1 + a, \ldots, x_n + a)$ indexed by $a \in \mathbb{R}$. The class of invariant tests is thus the class of tests that are of the form $\phi(\mathbf{x}) = \varphi(\mathbf{T}(\mathbf{x}))$, where $\mathbf{T}(\mathbf{x}) = (x_2 - x_1, \ldots, x_n - x_1)$ is the corresponding maximal invariant. Since the group of translations is a generating group for the null submodel, the maximal invariant, hence also all invariant tests, are distribution-free under the null hypothesis. In the induced model, the testing problem consists in testing

$$\mathcal{H}_0 : \mathbf{T} \sim f_0^{\mathbf{T}}(\mathbf{t}) = \int_{-\infty}^{\infty} f_0(z, t_1 + z, \ldots, t_{n-1} + z)\, dz$$

against

$$\mathcal{H}_1 : \mathbf{T} \sim g_0^{\mathbf{T}}(\mathbf{t}) = \int_{-\infty}^{\infty} g_0(z, t_1 + z, \ldots, t_{n-1} + z)\, dz$$

at level α. This is a problem involving a simple null hypothesis against a simple alternative hypothesis, for which the test defined by

$$\varphi_\alpha^*(\mathbf{t}) := \begin{cases} 1 & \text{if } g_0^{\mathbf{T}}(\mathbf{t}) > k f_0^{\mathbf{T}}(\mathbf{t}) \\ \gamma & \text{if } g_0^{\mathbf{T}}(\mathbf{t}) = k f_0^{\mathbf{T}}(\mathbf{t}) \\ 0 & \text{if } g_0^{\mathbf{T}}(\mathbf{t}) < k f_0^{\mathbf{T}}(\mathbf{t}), \end{cases}$$

where k and γ are such that $\mathrm{E}_{f_0^{\mathbf{T}}}[\varphi_\alpha^*(\mathbf{T})] = \alpha$, is (U)MP test at level α. The test

$$\phi_\alpha^*(\mathbf{x}) := \begin{cases} 1 & \text{if } g_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) > k f_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) \\ \gamma & \text{if } g_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) = k f_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) \\ 0 & \text{if } g_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) < k f_0^{\mathbf{T}}(\mathbf{T}(\mathbf{x})) \end{cases}$$

is therefore UMPI at level α for the original testing problem.

As a second example, consider the problem of testing $\mathcal{H}_0 : \sigma^2 \leq \sigma_0^2$ against $\mathcal{H}_1 : \sigma^2 > \sigma_0^2$ at level α when one observes $\mathbf{X} = (X_1, \ldots, X_n)$, where the X_i's are i.i.d. Gaussian with mean μ and variance σ^2 (the location parameter μ thus plays the role of a nuisance). This testing problem is invariant under the same group of translations as in the first example, so that the invariance principle leads to restricting to tests that are measurable with respect to the same maximal invariant $\mathbf{T}(\mathbf{x}) = (x_2 - x_1, \ldots, x_n - x_1)$. If $\mathbf{X} \sim \mathrm{P}_{\mu,\sigma^2}$, then $\mathbf{T} = \mathbf{T}(\mathbf{X})$ is multivariate normal with mean vector zero and covariance matrix $\sigma^2\mathbf{V}$, with

$$\mathbf{V} := \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 1 & & 1 \\ \vdots & & \ddots & \vdots \\ 1 & & \dots & 2 \end{pmatrix}.$$

Thus, the induced model is associated with the densities

$$f^{\mathbf{T}}_{\sigma^2}(\mathbf{t}) = \left(\frac{1}{2\pi\sigma^2}\right)^{(n-1)/2} \exp\left(-\frac{1}{2\pi\sigma^2}\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}\right), \qquad \sigma^2 > 0.$$

This is an exponential family with natural parameter $\theta = -1/(2\pi\sigma^2)$ and natural statistic $\mathbf{t}'\mathbf{V}^{-1}\mathbf{t}$. For the problem of testing $\mathcal{H}_0 : \sigma^2 \leq \sigma_0^2$ against $\mathcal{H}_1 : \sigma^2 > \sigma_0^2$ at level α in this induced model (which, letting $\theta_0 = -1/(2\pi\sigma_0^2)$ is equivalent to the problem of testing $\mathcal{H}_0 : \theta \leq \theta_0$ against $\mathcal{H}_1 : \theta > \theta_0$), the test defined by

$$\varphi^*_\alpha(\mathbf{t}) = \begin{cases} 1 & \text{if } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} > c_\alpha \\ \gamma_\alpha & \text{if } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} = c_\alpha \\ 0 & \text{if } \mathbf{t}'\mathbf{V}^{-1}\mathbf{t} < c_\alpha, \end{cases}$$

where c_α and γ_α are such that $\mathrm{E}_{\sigma_0^2}[\varphi^*_\alpha] = \alpha$, is UMP at level α. The resulting test

$$\phi^*_\alpha(\mathbf{x}) = \varphi^*_\alpha(\mathbf{T}(\mathbf{x})) = \begin{cases} 1 & \text{if } (\mathbf{T}(\mathbf{x}))'\mathbf{V}^{-1}\mathbf{T}(\mathbf{x}) > c_\alpha \\ \gamma_\alpha & \text{if } (\mathbf{T}(\mathbf{x}))'\mathbf{V}^{-1}\mathbf{T}(\mathbf{x}) = c_\alpha \\ 0 & \text{if } (\mathbf{T}(\mathbf{x}))'\mathbf{V}^{-1}\mathbf{T}(\mathbf{x}) < c_\alpha \end{cases}$$

is then UMPI at level α in the original model. Algebraic computations actually yield

$$\mathbf{V}^{-1} := \frac{1}{n}\begin{pmatrix} n-1 & -1 & \dots & -1 \\ -1 & n-1 & & -1 \\ \vdots & & \ddots & \vdots \\ -1 & & \dots & n-1 \end{pmatrix} \qquad \text{and} \quad (\mathbf{T}(\mathbf{x}))'\mathbf{V}^{-1}\mathbf{T}(\mathbf{x}) = ns^2$$

(with $s^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$), so that the test

$$\phi^*_\alpha(\mathbf{x}) = \begin{cases} 1 & \text{if } ns^2/\sigma_0^2 > \chi^2_{n-1,1-\alpha} \\ 0 & \text{if } ns^2/\sigma_0^2 \leq \chi^2_{n-1,1-\alpha}, \end{cases}$$

where $\chi^2_{n-1,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2_{n-1} distribution, is UMPI for the original problem. Note that this test is the one we had shown to be UMPU at level α for the same problem in the previous chapter.

8.5 Equivariance

Assume that the family $\mathcal{P}$ is invariant under $\mathcal{G},\circ$. For simplicity, assume that $\mathcal{P}$ is a parametric family with parameter $\boldsymbol{\theta}$, and consider the problem of estimating $\boldsymbol{\theta}$. If $\mathbf{X} \sim \mathrm{P}^{\mathbf{X}}_{\boldsymbol{\theta}}$ and $\mathbf{T}(\mathbf{X})$ is a "good" estimator of $\boldsymbol{\theta}$, then one may argue that, since $g(\mathbf{X}) \sim \mathrm{P}^{g(\mathbf{X})}_{\boldsymbol{\theta}} = \mathrm{P}^{\mathbf{X}}_{\bar{g}(\boldsymbol{\theta})}$, the relation of $\mathbf{T}(g(\mathbf{X}))$ to $\bar{g}(\boldsymbol{\theta})$ is the same as that of $\mathbf{T}(\mathbf{X})$ to $\boldsymbol{\theta}$, so that

$$\mathbf{T}(g(\mathbf{X})) \text{ should be a "good" estimator of } \bar{g}(\boldsymbol{\theta}). \tag{8.3}$$

This is, however, a "soft argument", and there is nothing mathematically compelling about (8.3). Now, another soft argument is: if $\mathbf{T}(\mathbf{X})$ qualifies as a "good" estimator of $\boldsymbol{\theta}$, that qualification should resist the transformations of $\bar{\mathcal{G}}$: more precisely,

$$\bar{g}(\mathbf{T}(\mathbf{X})) \text{ should be a "good" estimator of } \bar{g}(\boldsymbol{\theta}). \tag{8.4}$$

If both points of view (8.3) and (8.4) are to be reconciled, then $\mathbf{T}$ should satisfy

$$\mathbf{T}(g(\mathbf{X})) = \bar{g}(\mathbf{T}(\mathbf{X})) \quad \text{for all } g \in \mathcal{G}.$$

The following definition formalizes that property in a general context where the family $\mathcal{P}$ needs not be parametric.

Definition 5. *Let the model $(\mathcal{X}, \mathcal{A}, \mathcal{P})$, be invariant under the group $\mathcal{G},\circ$. An estimator $\mathbf{T} = \mathbf{T}(\mathbf{X})$ such that $\mathbf{T}(g(\mathbf{x})) = \bar{g}(\mathbf{T}(\mathbf{x}))$ for all $g \in \mathcal{G}$ and all $x \in \mathcal{X}$ is called* equivariant *(under $\mathcal{G},\circ$).*

Intuitively, an equivariant estimator has the same behavior, under the action $\mathcal{G}$, as the parameter it is estimating. If, however, (8.3) and (8.4) are to make sense, then the loss function $(\mathbf{t}, \boldsymbol{\theta}) \mapsto L_{\boldsymbol{\theta}}(\mathbf{t})$

(where $L_{\boldsymbol{\theta}}(\mathbf{t})$ is the loss incurred if the estimator takes value $\mathbf{t}$ when the true parameter value is $\boldsymbol{\theta}$) used in the definition of a "good" estimator should be compatible with the group structure of the model, and satisfy

$$L_{\bar{g}(\boldsymbol{\theta})}(\bar{g}(\mathbf{t})) = L_{\boldsymbol{\theta}}(\mathbf{t}) \quad \text{for all } g \in \mathcal{G} \text{ and } \boldsymbol{\theta} \in \Theta. \tag{8.5}$$

Definition 6. *A loss function* $(\mathbf{t}, \mathrm{P}) \mapsto L_{\mathrm{P}}(\mathbf{t})$ *(where* $L_{\mathrm{P}}(\mathbf{t})$ *is the loss incurred if the estimator takes value* $\mathbf{t}$ *when* $\mathbf{X}$ *has distribution* P*) is called* invariant *if*

$$L_{\bar{g}(\mathrm{P})}(\mathbf{T}(g(\mathbf{x}))) = L_{\mathrm{P}}(\mathbf{T}(\mathbf{x})) \quad \textit{for all } g \in \mathcal{G},\ \mathbf{x} \in \mathcal{X} \textit{ and } \mathrm{P} \in \mathcal{P}. \tag{8.6}$$

Note that, in the parametric case, (8.6) applied to equivariant estimators $\mathbf{T}$ reduces to (8.5).

If equivariance is considered a reasonable property, then one may adhere to the *Principle of equivariance*, which consists (in the context described in this chapter) in restricting to equivariant estimators; just as the *Principle of unbiasedness*, it is a *statistical principle*, that is, a rule one may decide or not to adopt. As we shall see, however, the benefits of equivariance are that optimal equivariant estimators typically exist in models that are *generated* by a group $\mathcal{G}, \circ$. This is actually based on the following property (parametric notation is used for simplicity).

Proposition 2. *The risk* $R^{\mathbf{T}}_{\boldsymbol{\theta}}$ *of an equivariant estimator* $\mathbf{T}$*, for an invariant loss function* $L_{\boldsymbol{\theta}}(\mathbf{t})$*, is constant along the orbits of* $\bar{\mathcal{G}}, \circ$.

Proof. First note that, for any measurable function Ψ which is $\mathrm{P}_{\boldsymbol{\theta}}$-integrable for all $\boldsymbol{\theta}$,

$$\mathrm{E}_{\bar{g}(\boldsymbol{\theta})}[\Psi(\mathbf{X})] = \int \Psi(\mathbf{x}) d\mathrm{P}^{\mathbf{X}}_{\bar{g}(\boldsymbol{\theta})}(\mathbf{x}) = \int \Psi(\mathbf{x}) d\mathrm{P}^{g\mathbf{X}}_{\boldsymbol{\theta}}(\mathbf{x}) = \mathrm{E}_{\boldsymbol{\theta}}[\Psi(g(\mathbf{X}))]. \tag{8.7}$$

Now, $R^{\mathbf{T}}_{\bar{g}(\boldsymbol{\theta})} = \mathrm{E}_{\bar{g}(\boldsymbol{\theta})}[L_{\bar{g}(\boldsymbol{\theta})}(\mathbf{T}(\mathbf{X}))]$. Applying (8.7) with $\Psi = L_{\bar{g}(\boldsymbol{\theta})}$ yields, since $\mathbf{T}$ is equivariant and the loss L invariant,

$$R^{\mathbf{T}}_{\bar{g}(\boldsymbol{\theta})} = \mathrm{E}_{\bar{g}(\boldsymbol{\theta})}[L_{\bar{g}(\boldsymbol{\theta})}(\mathbf{T}(\mathbf{X}))] = \mathrm{E}_{\boldsymbol{\theta}}[L_{\bar{g}(\boldsymbol{\theta})}(\mathbf{T}(g(\mathbf{X})))] = \mathrm{E}_{\boldsymbol{\theta}}[L_{\bar{g}(\boldsymbol{\theta})}(\bar{g}(\mathbf{T}(\mathbf{X})))] = \mathrm{E}_{\boldsymbol{\theta}}[L_{\boldsymbol{\theta}}(\mathbf{T}(\mathbf{X}))] = R^{\mathbf{T}}_{\boldsymbol{\theta}}.$$

The result follows. □

Corollary 2. *If* $\mathcal{G}, \circ$ *is a generating group and the loss function is invariant, the risk* $R^{\mathbf{T}}_{\boldsymbol{\theta}}$ *of any equivariant estimate* $\mathbf{T}$ *is constant (its value does not depend on* $\boldsymbol{\theta}$*).*

This corollary has the very important consequence that, when the model is generated by a group $\mathcal{G}$, then two equivariant estimators ($\mathbf{T}_1$ and $\mathbf{T}_2$, say) are always comparable: either

$$R_{\boldsymbol{\theta}}^{\mathbf{T}_1} \leq R_{\boldsymbol{\theta}}^{\mathbf{T}_2} \text{ for all } \boldsymbol{\theta}, \text{ and } \mathbf{T}_1 \succeq \mathbf{T}_2$$

or

$$R_{\boldsymbol{\theta}}^{\mathbf{T}_1} \geq R_{\boldsymbol{\theta}}^{\mathbf{T}_2} \text{ for all } \boldsymbol{\theta}, \text{ and } \mathbf{T}_1 \preceq \mathbf{T}_2.$$

Uniformly minimum risk equivariant estimators (UMRE) thus are very likely to exist for any invariant loss function.

8.6 UMRE estimates for location

Consider an observed n-tuple $\mathbf{X} = (X_1, \ldots, X_n)$ in a location family. Let $\tilde{\theta}(\mathbf{x})$ be equivariant (such $\tilde{\theta}$ exists: take for instance $\tilde{\theta}(\mathbf{x}) = x_1$). Denote by $\hat{\theta}(\mathbf{x})$ any other equivariant estimator (such $\hat{\theta}$ exists: take, for instance, $\hat{\theta}(\mathbf{x}) = x_2$). Equivariance of $\tilde{\theta}$ and $\hat{\theta}$ implies that, for any $a \in \mathbb{R}$,

$$\begin{aligned} \hat{\theta}(\mathbf{x} + a\mathbf{1}) &= \hat{\theta}(\mathbf{x}) + a, \\ \tilde{\theta}(\mathbf{x} + a\mathbf{1}) &= \tilde{\theta}(\mathbf{x}) + a, \end{aligned}$$

hence

$$\hat{\theta}(\mathbf{x} + a\mathbf{1}) - \tilde{\theta}(\mathbf{x} + a\mathbf{1}) = \hat{\theta}(\mathbf{x}) - \tilde{\theta}(\mathbf{x}).$$

It follows that $\hat{\theta} - \tilde{\theta}$ is invariant; therefore, it is measurable with respect to the maximal invariant $\mathbf{T} = \mathbf{T}(\mathbf{X}) := (X_2 - X_1, X_3 - X_1, \ldots, X_n - X_1)$ (equivalently, $\hat{\theta} - \tilde{\theta}$ is $\mathcal{A}_{\mathcal{G}}$-measurable). Since this holds for any equivariant $\tilde{\theta}$ and $\hat{\theta}$, the class of all equivariant estimators can be described as

$$\left\{ \tilde{\theta}(\mathbf{X}) + \psi(\mathbf{T}) : \psi \text{ measurable} \right\},$$

where $\tilde{\theta}$ denotes an arbitrary initial equivariant estimator. Thus, a UMRE estimator, if it exists, is of the form

$$\theta^*(\mathbf{X}) = \tilde{\theta}(\mathbf{X}) + \psi^*(\mathbf{T})$$

where ψ^* (uniformly in θ) minimizes (over all measurable functions) the risk: for any $\theta \in \Theta$ and measurable ψ,

$$\mathrm{E}_\theta\big[L_\theta\big(\tilde{\theta} + \psi^*(\mathbf{T})\big)\big] \leq \mathrm{E}_\theta\big[L_\theta\big(\tilde{\theta} + \psi(\mathbf{T})\big)\big]. \tag{8.8}$$

Such ψ^* typically depends on the loss function adopted.

8.6.1 Quadratic loss

Under quadratic loss $L_\theta(t) = (t-\theta)^2$, the condition (8.8) takes the form

$$\mathrm{E}_\theta\big[\big(\tilde\theta + \psi^*(\mathbf{T}) - \theta\big)^2\big] \le \mathrm{E}_\theta\big[\big(\tilde\theta + \psi(\mathbf{T}) - \theta\big)^2\big] \tag{8.9}$$

for all $\theta \in \Theta$ and all measurable ψ, or equivalently,

$$\mathrm{E}_\theta\big[\big((\theta - \tilde\theta) - \psi^*(\mathbf{T})\big)^2\big] \le \mathrm{E}_\theta\big[\big((\theta - \tilde\theta) - \psi(\mathbf{T})\big)^2\big] \tag{8.10}$$

for all $\theta \in \Theta$ and all measurable ψ. As seen when covering conditional expectations in Chapter 2, this condition is satisfied for

$$\psi^*(\mathbf{T}) = \mathrm{E}_\theta\big[\theta - \tilde\theta|\mathbf{T}\big] = -\mathrm{E}_\theta\big[\tilde\theta - \theta|\mathbf{T}\big].$$

Hence, the UMRE for quadratic loss is

$$\begin{aligned}\theta^* &= \tilde\theta - \mathrm{E}_\theta[\tilde\theta - \theta|\mathbf{T}]\\ &= \tilde\theta - \mathrm{E}_0[\tilde\theta|\mathbf{T}],\end{aligned}$$

since (due to the equivariance of $\tilde\theta(\mathbf{X})$ and invariance of $\mathbf{T}$)

$$(\tilde\theta(\mathbf{X}) - \theta, \mathbf{T}(\mathbf{X})) = (\tilde\theta(\mathbf{X} - \theta\mathbf{1}), \mathbf{T}\,(\mathbf{X} - \theta\mathbf{1})),$$

which, under $\mathbf{X} \sim \mathrm{P}_\theta$ has the same distribution as $(\tilde\theta(\mathbf{X}), \mathbf{T}(\mathbf{X}))$ under $\mathbf{X} \sim \mathrm{P}_0$; formally, $\mathrm{P}_\theta^{(\tilde\theta(\mathbf{X})-\theta,\mathbf{T}(\mathbf{X}))} = \mathrm{P}_\theta^{(\tilde\theta(\mathbf{X}-\theta\mathbf{1}),\mathbf{T}(\mathbf{X}-\theta\mathbf{1}))} = \mathrm{P}_0^{(\tilde\theta(\mathbf{X}),\mathbf{T}(\mathbf{X}))}$. It follows that θ^* is a statistic that does not depend on θ, and that it is also *essentially unique*.

<u>Remark 1</u>: θ^* is unbiased: for any $\theta \in \Theta$, we indeed have

$$\mathrm{E}_\theta[\theta^*] = \mathrm{E}_\theta[\tilde\theta - \mathrm{E}_\theta[\tilde\theta - \theta|\mathbf{T}]] = \mathrm{E}_\theta[\tilde\theta] - \mathrm{E}_\theta[\tilde\theta - \theta] = \theta$$

(actually, θ^* is obtained by subtracting from $\tilde\theta$ its conditional bias: therefore, θ^* is conditionally unbiased, hence unconditionally unbiased). The corresponding risk is thus $\mathrm{R}_\theta^{\theta^*} = \mathrm{Var}_\theta[\theta^*]$.

Remark 2: In the Gaussian sampling model $\mathbf{X} = (X_1, \dots, X_n)$, where the X_i's are i.i.d. $\mathcal{N}(\mu, \sigma_0^2)$ with $\mu \in \mathbb{R}$, we have $\theta^* = \bar{X}$ almost surely. Indeed, denoting by $\mathrm{R}_\theta^{\bar{X}}$ the quadratic risk for $\bar{X}$, we have that, under Gaussian density,

$$\mathrm{R}_\theta^{\bar{X}} \geq \mathrm{R}_\theta^{\theta^*} \ \forall \theta \in \Theta, \text{ since } \bar{X} \text{ is equivariant and } \theta^* \text{ is UMRE} \tag{8.11a}$$

$$\mathrm{R}_\theta^{\bar{X}} \leq \mathrm{R}_\theta^{\theta^*} \ \forall \theta \in \Theta, \text{ since } \bar{X} \text{ is UMVU and } \theta^* \text{ is unbiased.} \tag{8.11b}$$

Hence, $\mathrm{R}_\theta^{\bar{X}} = \mathrm{R}_\theta^{\theta^*}$ $\forall \theta \in \Theta$, $\bar{X}$ is also UMRE, and $\bar{X} = \theta^*$ P_θ-almost surely, in view of the P_θ-almost sure unicity of θ^*. Since nothing here depends on σ_0^2, the same remark holds for the case of an unspecified variance σ^2.

Remark 3: Note that (8.11) in Remark 2 always holds:

$$\mathrm{Var}_\theta[\theta^*] = \mathrm{R}_\theta^{\theta^*} \leq \mathrm{R}_\theta^{\bar{X}} = \mathrm{Var}_\theta[\bar{X}] = \frac{\sigma^2}{n}$$

(assuming finite moments of order two), whereas, in the Gaussian case,

$$\mathrm{Var}_\theta[\theta^*] = \mathrm{Var}_\theta[\bar{X}] = \frac{\sigma^2}{n} \cdot$$

It follows that the Gaussian case is "least favorable" for equivariant estimation of location under quadratic loss.

Remark 4: *Pitman form.* Choose $\tilde{\theta} = X_1$. Then,

$$\begin{aligned}
\theta^* &= X_1 - \mathrm{E}_0\big[X_1 | \overbrace{X_2 - X_1}^{Y_2}, \ldots, \overbrace{X_n - X_1}^{Y_n}\big] \\
&= X_1 - \frac{\int u f_0(u, Y_2 + u, \ldots, Y_n + u)\,\mathrm{d}u}{\int f_0(u, Y_2 + u, \ldots, Y_n + u)\,\mathrm{d}u} \\
&= X_1 - \frac{\int u f_0(u, X_2 - X_1 + u, \ldots, X_n - X_1 + u)\,\mathrm{d}u}{\int f_0(u, X_2 - X_1 + u, \ldots, X_n - X_1 + u)\,\mathrm{d}u} \\
&= X_1 - \frac{\int (X_1 - t) f_0(X_1 - t, X_2 - t, \ldots, X_n - t)\,\mathrm{d}t}{\int f_0(X_1 - t, X_2 - t, \ldots, X_n - t)\,\mathrm{d}t} \\
&= \cancel{X_1} - \cancel{X_1} + \frac{\int t f_0(X_1 - t, X_2 - t, \ldots, X_n - t)\,\mathrm{d}t}{\int f_0(X_1 - t, X_2 - t, \ldots, X_n - t)\,\mathrm{d}t} \\
&= \frac{\int t f_t(X_1, X_2, \ldots, X_n)\,\mathrm{d}t}{\int f_t(X_1, X_2, \ldots, X_n)\,\mathrm{d}t},
\end{aligned}$$

where we let $u = X_1 - t$ in order to obtain a more symmetric expression.

Remark 5: Still for quadratic loss, if a UMVU estimator exists and is equivariant, then it is UMRE (compare with Remark 2).

8.6.2 Absolute deviation loss

For the absolute deviation loss is given by $L_\theta(t) = |t - \theta|$, Condition (8.8) takes the form

$$\mathrm{E}_\theta[|\tilde{\theta} - \psi^*(\mathbf{T}) - \theta|] \leq \mathrm{E}_\theta[|\tilde{\theta} - \psi(\mathbf{T}) - \theta|]$$

for any measurable ψ. This condition is satisfied for $\psi^*(\mathbf{T}) = -\mathrm{Med}_\theta[\tilde{\theta} - \theta | \mathbf{T}]$, where $\mathrm{Med}_\theta[\tilde{\theta} - \theta | \mathbf{T}]$ denotes the conditional median of $\tilde{\theta} - \theta$ (conditional on $\mathbf{T}$). This readily follows from the fact that, for any $(\mathcal{X}, \mathcal{A})$-measurable ζ and sub-sigma field $\mathcal{A}_0 \subseteq \mathcal{A}$,

$$\arg \min_{\mu, \mathcal{A}_0\text{-measurable}} \mathrm{E}\left[|\zeta - \mu|\right] = \mathrm{Med}\left[\zeta | \mathcal{A}_0\right].$$

Hence, the UMRE for the least absolute deviation (LAD) estimator is

$$\begin{aligned} \theta^* &= \tilde{\theta} - \mathrm{Med}_\theta[\tilde{\theta} - \theta | \mathbf{T}] \\ &= \tilde{\theta} - \mathrm{Med}_0[\tilde{\theta} | \mathbf{T}], \end{aligned}$$

with the same reasoning as above.