5 Equivariance

Assuming that the family of distributions, containing that unknown distribution that data are observed from, has the property of being invariant or equivariant under some transformation, it is natural to demand that also the estimator satisfies the same invariant/equivariant property.

5.1 The principle of equivariance

Let $\mathcal{P} = \{P_\theta : \theta \in \Omega\}$ be a family of distributions. Let $\mathcal{G} = \{g\}$ be a class of transformations of the sample space, i.e. $g : \mathcal{X} \mapsto \mathcal{X}$, that is a group under composition.

Definition 8 *If* (i) *when* $X \sim P_\theta$ *and*

$$X' = gX \quad \sim \quad P_{\theta'} \in \mathcal{P},$$

for each $g \in \mathcal{G}$*, for some element* θ' *in* Ω *and* (ii) *if for each fixed* g *as* θ *traverses* Ω*, so does* $g\theta$*, then* $\mathcal{P}$ *is called invariant under* $\mathcal{G}$ *.* □

Example 19 *Let* F *be a fixed distribution on* $\mathbb{R}^n$*,* $\mathcal{P} = \{F(x - \theta) : \theta \in \mathbb{R}\}$ *and with* $g_a(x) = x + a$*,* $\mathcal{G} = \{g_a : a \in \mathbb{R}\}$*. Then* $\mathcal{P}$ *is invariant under* $\mathcal{G}$*.* □

Assume $\mathcal{G}$ is a group of transformations that leave $\mathcal{P}$ invariant. Then the map

$$g : P_\theta \sim X \quad \mapsto \quad gX \sim P_{\theta'}$$

induces a map $\bar{g}$ on Ω as

$$\bar{g} : \Omega \ni \theta \quad \mapsto \quad \theta' \in \Omega,$$

i.e. $\theta' = \bar{g}\theta$.

It is easy to see that if $\{P_\theta\}$ are distinct (i.e. different θ's give rise to different P_θ') and $\mathcal{G}$ is a group then $\bar{\mathcal{G}} = \{\bar{g}\}$ is a group.

Example 20 *(ctd.) In the previous example, with* $g_a(x) = x + a$ *and* $\Omega = \mathbb{R} = \{\theta\}$*, the induced maps on* Ω *are given by*

$$g_a : P_\theta \sim X \quad \mapsto \quad X + a \sim P_{\theta+a} = P_{\theta'},$$

so that $\theta' = \theta + a$*, i.e.* $\theta' = \bar{g}\theta$ *with*

$$\bar{g} : \theta \quad \mapsto \quad \theta + a.$$

□

The groups $\mathcal{G}, \bar{\mathcal{G}}$ are related via: For every measurable set A

$$P_\theta(gX \in A) \;=\; P_{\bar{g}\theta}(X \in A),$$

since $X \sim P_\theta$ implies $gX \sim P_{\bar{g}\theta}$, or equivalently

$$P_\theta(g^{-1}(A)) \;=\; P_{\bar{g}\theta}(A).$$

Now assume the set of probabilities $\mathcal{P}$ are invariant under the transformations $\mathcal{G}, \bar{\mathcal{G}}$, and assume we want to estimate the estimand $h(\theta)$, for some function h. What forms on h are possible?

Example 21 *Let f be a fixed density on $\mathbb{R}^{n+m}$ and assume $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ are jointly distributed according to a density in the location model family*

$$\mathcal{P} \;=\; \{f(x-\xi, y-\eta) : (\xi, \eta) \in \mathbb{R}^2\}.$$

Let

$$\begin{aligned} g_{a,b} : \mathbb{R}^{n+m} \ni (x, y) &\;\mapsto\; (x+a, y+b) \in \mathbb{R}^{n+m}, \\ \bar{g}_{a,b} : \mathbb{R}^2 \ni (\xi, \eta) &\;\mapsto\; (\xi+a, \eta+b) \in \mathbb{R}^2, \end{aligned}$$

and $\mathcal{G} = \{g_{a,b} : (a,b) \in \mathbb{R}^2\}$ and $\bar{\mathcal{G}} = \{\bar{g}_{a,b} : (a,b) \in \mathbb{R}^2\}$. Then $\mathcal{P}$ is invariant under the groups of transformations $\mathcal{G}, \bar{\mathcal{G}}$.

(i): Assume we want to estimate

$$h(\xi, \eta) \;=\; \eta - \xi.$$

Under the transformations $g_{a,b}, \bar{g}_{a,b}$ the estimand is transformed to

$$h(\xi', \eta') \;=\; h(\xi, \eta) + (b-a).$$

A sensible estimator should give values that are transformed from d (when based on the random variables X, Y) to $d' = d + (b-a)$ (when based on the random variables $(X', Y') = (X+a, Y+a)$). (This is really the defining property for an equivariant estimator, cf. the sequel.) The estimation problem can be labeled invariant if the loss function satisfies $L((\xi', \eta'), d + (b-a)) = L((\xi, \eta), d)$. Such loss functions exist: The condition is equivalent to the loss function being of the form $L((\xi, \eta), d)) = \rho(h(\xi, \eta) - d)$ for some function ρ.

(ii): Assume we want to estimate

$$h(\xi, \eta) \;=\; \xi^2 + \eta^2.$$

Under the transformations $g_{a,b}, \bar{g}_{a,b}$ the estimand is transformed to

$$\begin{aligned} h(\xi', \eta') &\;=\; (\xi+a)^2 + (\eta+b)2 \\ &\;=\; h(\xi, \eta) + 2(\xi a + \eta b) + a^2 + b^2 \\ &\;\neq\; h(\xi, \eta) + \kappa(a, b), \end{aligned}$$

for any function κ that depends on only (a, b). But a sensible estimator should give values that are transformed from d (when based on (X, Y)) to $d' = d + 2(\xi a + \eta b) + a^2 + b^2$ (when based on (X', Y')) to keep the principle of invariance, or equivalently put to be an equivariant estimator. However the value d' is transformed via a transformation that depends on the unknowns (ξ, η), and this is a not a realizable estimator (it depends on the unknown parameters). □

To keep the principle of invariance, and the estimation model invariant, we thus need to have an estimand that after the transformation depends on θ only through $h(\theta)$. This means that $h(\bar{g}\theta)$ should depend on θ only through $h(\theta)$. Now assume that this holds. Then the map $\bar{g}$ generates a map on value space $\mathcal{H} = \{h(\theta) : \theta \in \Omega\}$ of the estimand as

$$\begin{aligned} \bar{g} : \theta &\mapsto \theta', \\ g^* : \mathcal{H} \ni h(\theta) &\mapsto h(\bar{g}\theta) \in \mathcal{H}, \end{aligned}$$

so that $g^* h(\theta) = h(\bar{g}\theta)$. Note that the condition that $h(\bar{g}\theta)$ should depend on θ only through $h(\theta)$ makes g^* a well defined map. Let $G^* = \{g^*\}$ denote the set of such transformations. It is easy to see that if $\bar{\mathcal{G}}$ is a group then $\mathcal{G}^*$ is a group.

Now assume that $\mathcal{G}, \bar{\mathcal{G}}, \mathcal{G}^*$ leave $\mathcal{P}$ invariant and the estimand is $h(\theta)$. The loss function is called invariant if

$$L(\bar{g}\theta, g^* d) = L(\theta, d).$$

Such a loss function gives the same loss for an estimator value d based on X as far the transformed estimator value $g^* d$ (which is the same transformation as the estimand undergoes) based on the transformed gX.

If $\mathcal{G}, \bar{\mathcal{G}}, \mathcal{G}^*$ leave $\mathcal{P}$ invariant and L is an invariant loss function, the estimation problem is called invariant.

It is now clear how to define an equivariant estimator.

Definition 9 *An estimator $\delta(X)$ for an invariant estimation problem is called equivariant if*

$$\delta(gX) = g^* \delta(X),$$

for every $g \in \mathcal{G}, g^* \in \mathcal{G}^*$.

5.2 Location equivariance

Recall that a location family of distributions was given by

$$\mathcal{P} = \{f(x - \xi) : \xi \in \Omega\}.$$

In this section we will treat that case that the sample space is $\Xi = \mathbb{R}^n$ and that $\Omega = \mathbb{R}$, but other setups are possible, cf. Examples 6 and 7.

Definition 10 *(Location invariance) Let $\mathcal{P}$ be a location family, and $L(\xi, d)$ a loss function. The loss functions is called location invariant if*

$$L(\xi + a, d + a) \;=\; L(\xi, d),$$

for every $\xi \in \Omega, d, a \in \mathbb{R}$. If L is location invariant the estimation problem is called location invarinant.

Note that by construnction $f_{\xi+a}(x+a) = f_\xi(x)$. Note also that for the location model the group operations are given by

$$g_a(\bar{g}_a) : z \mapsto z + a,$$

with $a \in \mathbb{R}$.

Definition 11 *(Location equivariants estimator) If $\mathcal{P}, L$ is a location invariant estimation problem and δ is an estimator. Then δ is called location equivariant if*

$$\delta(x + a) \;=\; \delta(x) + a.$$

Note that this is also consistent with our definition of equivariant estimators since for location families the group operation is given by $g_a^* : d \mapsto d + a$.

Theorem 4 *Assume that $(X, \mathcal{P}, L)$ is a location invariant problem, and δ is an equivariant estimator. Then the bias, variance and risk of $\delta(X)$ are all constant (i.e. they do not depend on ξ).*

Proof. (i). The bias is

$$\begin{aligned} E_\xi(\delta(X)) - \xi &= E_0(\delta(X + \xi)) - \xi \\ &= E_0(\delta(X)) + \xi - \xi \\ &= E_0(\delta(X)), \end{aligned}$$

where the first equality is by the invariance of the family and the second is the linearity of the expectation. Thus the bias does not depend on ξ.

(ii): The variance is

$$\begin{aligned} Var_\xi(\delta(X)) &= E_\xi(\delta(X)^2) - (E_\xi(\delta(X)))^2 \\ &= E_0(\delta(X + \xi)^2) - (E_0(\delta(X + \xi))^2 \\ &= E_0(\delta^2(X) + 2\delta(X)\xi + \xi^2) - E_0(\delta(X))^2 - 2\xi E_0(\delta(X)) - \xi^2 \\ &= Var_0(\delta(X)), \end{aligned}$$

where the second equality follows by the invariance of the family and the third follows by the equivariance of the estimator. Thus the variance is independent of ξ.

(iii): The risk is

$$
\begin{aligned}
E_\xi(L(\xi, \delta(X))) &= E_0(L(\xi, \delta(X + \xi))) \\
&= E_0(L(\xi, \delta(X) + \xi)) \\
&= E_0(L(0, \delta(X))),
\end{aligned}
$$

where the first equality follows by the invariance of the family, the second by the equivariance of the estimator, and the third by the invariance of the loss function. Thus the risk is independent of ξ.

□

We next give a characterization of the set of equivariant estimators that is useful.

Lemma 3 *Let δ_0 be a fixed equivariant estimator. The set of (location) equivariant estimators is given by*

$$
\Delta = \{\delta = \delta_0 + u : u(x) = u(x + a), \textit{ for all } x \in \mathcal{X}, a \in \mathbf{R}\}.
$$

Proof. Assume first that δ_0 is fixed equivariant estimator, and u and invariant estimator , i.e. $u(x + a) = u(x)$ for all x, a. Define $\delta = \delta_0 + u$. Then

$$
\begin{aligned}
\delta(x + a) &= \delta_0(x + a) + u(x + a) \\
&= \delta_0(x) + a + u(x) \\
&= \delta(x) + a,
\end{aligned}
$$

i.e. δ is equivariant.

Assume instead that δ_0 is a fixed equivariant estimator and let δ be an arbitrary equivariant estimator. Define $u = \delta - \delta_0$. Then u is invariant:

$$
\begin{aligned}
u(x + a) &= \delta(x + a) - \delta_0(x + a) \\
&= \delta(x) + a - \delta_0(x) - a \\
&= \delta(x) - \delta_0(x) \\
&= u(x),
\end{aligned}
$$

and $\delta = \delta_0 + u$. □

Note that this means that we get all equivariant estimators by taking one fixed such δ_0 and add an estimator u as above. Estimators u such as above we can call invariant. Thus the totality of all equivariant estimators is obtained by taking one fixed such estimator and going through (adding) the totality of invariant estimators.

Next we give a characterization of the invariant estimators u in the above characterization of Δ. Let $\mathcal{X} = \mathbb{R}^n$ and $\Omega = \mathbb{R}$. Define the set

$$
\mathcal{U} = \{u : u(x + a) = u(x), x \in \mathcal{X}, a \in \Omega\}
$$

of invariant (functions) estimators.

Lemma 4 *Under the above assumptions*

$$\mathcal{U} = \{u(x_1, \ldots, x_n) = h(x_1 - x_n, \ldots, x_{n-1} - x_n) : h \textit{ function on } \mathbb{R}^{n-1}\}.$$

Proof. ($\supset$) Assume that $u = h$ for a function $h : \mathbb{R}^{n-1} \to \mathbb{R}$. Then

$$\begin{aligned} u(x + a) &= h(x_1 + a - x_n - a, \ldots, x_{n-1} + a - x_n - a) \\ &= h(x_1 - x_n, \ldots, x_{n-1} - x_n) \\ &= u(x), \end{aligned}$$

and thus u is invariant.

($\subset$) Converseley, assume instead u is invariant so $u(x + a) = u(x)$ for all x, a. Define the function $h : \mathbb{R}^{n-1} \to \mathbb{R}$ by $h(x_1, \ldots, x_{n-1}) = u(x_1, \ldots, x_{n-1}, 0)$. Then by the invariance of u

$$\begin{aligned} u(x_1, \ldots, x_n) &= u(x_1 - x_n, x_2 - x_n, \ldots, x_n - x_n) \\ &= h(x_1 - x_n, \ldots, x_{n-1} - x_n). \end{aligned}$$

$\square$

In particular when $n = 1$, u is invariant if and only if u is a constant (function of $x_1 - x_1 = 0$).

If we combine the previous we get the following characterization of the set Δ of equivariant estimators.

Theorem 5 *Let δ_0 be a fixed equivariant estimator. Under the above assumptions*

$$\Delta = \{\delta = \delta_0 - v : v \textit{ function defined on } \mathbb{R}^{n-1}\}.$$

Definition 12 *Assume we have a location equivariant inference problem. If*

$$\hat{\delta} = \operatorname{argmin}_{\delta \in \Delta} R(\xi, \delta)$$

exists it is called the Minimum Risk Equivariant (MRE) estimator of ξ.

$\square$

Recall that the risk for equivariant estimators is independent of the parameter ξ, and thus if the MRE estimator exists it mimimizes the risk *uniformly* over all parameter values ξ. Note also that if it exists we can define the MRE as

$$\begin{aligned} \hat{\delta} &= \operatorname{argmin}_{\delta \in \Delta} R(0, \delta) \\ &= \operatorname{argmin}_{\delta \in \Delta} E_0(L(0, \delta(X))). \end{aligned}$$

Assuming that L is invariant, one could ask what forms of L are possible?

Lemma 5 *The set of invariant loss function is*

$$\{L(\xi, d) = \rho(d - \xi) : \rho \text{ function } \mathbb{R} \to \mathbb{R}^+, \rho(0) = 0\}.$$

Proof. Assume first that L is invariant, i.e. that for all a

$$L(\xi + a, d + a) \;=\; L(\xi, d).$$

Define $\rho(u) = L(0, u)$. Put $a = -\xi$ to get $L(\xi, d) = L(0, d - \xi) = \rho(d - \xi)$.

Conversely, assume $L(\xi, d) = \rho(d - \xi)$. Then

$$\begin{aligned} L(\xi + a, d + a) &= \rho(d + a - \xi - a) \\ &= \rho(d - a) \\ &= L(\xi, d), \end{aligned}$$

so that L is invariant. □

The next result gives the MRE estimator for location families.

Theorem 6 *Assume $X = (X_1, \ldots, X_n)$ is distributed according to a location family and let $Y = (Y_1, \ldots, Y_{n-1})$, with $Y_i = X_i - X_n$, for $i = 1, n-1$. Assume there exists a fixed equivariant estimator δ_0 of ξ, with finite risk. If*

$$\hat{v}(y) \;=\; \operatorname{argmin}_{\{v:\mathbb{R}^{n-1}\to\mathbb{R}\}} E_0\left(\rho(\delta_0(X) - v(y)) | y\right)$$

exists for each y, then

$$\hat{\delta}(x) \;=\; \delta_0(x) - \hat{v}(y),$$

is MRE.

Proof. By definition the MRE is given by

$$\begin{aligned} \hat{\delta} &= \operatorname{argmin}_{\delta\in\Delta} E_0\left(L(0, \delta(X))\right) \\ &= \delta_0 - \operatorname{argmin}_{\{v:\mathbb{R}^{n-1}\to\mathbb{R}\}} E_0\left(L(0, \delta_0(X) - v(Y))\right) \\ &= \delta_0 - \operatorname{argmin}_{\{v:\mathbb{R}^{n-1}\to\mathbb{R}\}} E_0\left(\rho(\delta_0(X) - v(Y))\right) \\ &= \delta_0 - \operatorname{argmin}_{\{v:\mathbb{R}^{n-1}\to\mathbb{R}\}} \int E_0\left(\rho(\delta_0(X) - v(Y)) |\, y\right) dP(y), \end{aligned}$$

where the second equality follows by the characterization of the set of equivariant estimators Δ in Theorem 5, and the third by the characterization of the invariant loss functions in Lemma 5. Since the integrand in the last line is non-negative and we integrate with respect to a probability measure dP, it follows that the integral is minimized by minimizing each integrand, which shows the statement of the theorem. □

We get following two special cases.

Corollary 2 *Under the assumptions in Theorem 6, we have the following two MRE estimators:*

(i) If $\rho(u) = u^2$ is the quadratic loss function then

$$\hat{\delta}(x) \;=\; \delta_0(x) - E_0(\delta_0(X)|y).$$

(ii) If $\rho(u) = |u|$ then

$$\hat{\delta}(x) \;=\; \delta_0(x) - \mathrm{med}(\delta_0(x)|y),$$

where $\mathrm{med}(\delta_0(X)|y)$ *is the conditional median of $\delta_0(X)$ given $Y = y$.*

Proof. (i) $E_0\left((\delta_0(X) - v(y))^2|y\right)$ is minimized by $\hat{v}(y) = E(\delta_0(X)|y)$. (ii) $E_0\left(|\delta_0(X) - v(y)||y\right)$ is minimized by the conditional median of $\delta_0(X)$ given $Y = y$. □

Example 22 *Assume we have one observation, i.e. $n = 1$. Then since an arbitrary equivariant estimator can be written as*

$$\begin{aligned}\delta(x) &= \delta_0(x) - v(x - x)\\ &= \delta_0(x) + c,\end{aligned}$$

for a fixed equivariant estimator $\delta_0(x)$ and arbitrary constact c, and since $\delta_0(x) = x$ is eqivariant, it follows that

$$\delta(x) \;=\; x + c$$

are the only equivariant estimators.

To find the MRE estimator we need to find

$$\hat{v} \;=\; \mathrm{argmin}_{v\in\mathbb{R}} E_0(\rho(x - v)).$$

If ρ is convex this is always possible, and $\hat{v}$ is unique if ρ is strictly convex (cases (i),(ii) below).

(i) If $\rho(u) = u^2$ then

$$\hat{v} \;=\; E_0(X)$$

is the MRE.

(ii) If $\rho(u) = |u|$ then

$$\hat{v} \;=\; \mathrm{med}(X)$$

is the MRE.

(iii) If $\rho(u) = 1\{|u| > k\}$ then minimizing $E_0(\rho(x - v)) = P_0(|x - v| > k)$ is the same as maximizing $P(|X - v| \leq k)$. The outcome of this optimization depends on the form of the distribution of X. a) Assume that F has a density f and that it is

symmetric around 0 and unimodal. Then $\hat{v} = 0$ *and thus the MRE is* $\hat{\delta}(x) = x - 0 = x$*. b) Assume instead that* f *is symmetric around zero and U-shaped, with support on* $[-c, c]$*. Then* $\hat{v}_1 = c - k$ *and* $\hat{v}_2 = k - c$ *are both minimizers and thus* $\hat{\delta}_1(x) = x - c - k$ *and* $\hat{\delta}_2(x) = x + c - k$ *are both MRE estimators. (Note that* ρ *is not strictly convex in this case.)*

□

Example 23 *Assume* $x_1, \ldots, x_n$ *are i.i.d.* $N(\xi, \sigma^2)$ *with* σ^2 *known. Let* $\delta_0 = \bar{x}$*. Then* δ_0 *is equivariant, and it is a complete sufficient statistic (it is the* T *in an exponential family). Also* $Y = (X_1, -X_n, \ldots, X_{n-1} - X_n)$ *is ancillary (it has a distribution that does not depend on* ξ*). Basu's theorem then says that* Y *is independent of* $\bar{X}$*. Thus*

$$\begin{aligned}\hat{v}(y) &= \operatorname{argmin}_v E_0(\rho(\delta_0(X) - v(y))|Y = y) \\ &= \operatorname{argmin}_v E_0(\rho(\delta_0(X) - v(y)))\end{aligned}$$

which is a constant $\hat{v}$ *(i.e. does not depend on* y*). Thus* $\bar{x} - \hat{v}$ *is the MRE. If* ρ *is convex and even, then since the distribution of* $\bar{X}$ *is symmetric around 0 under* E_0 *clearly* $\hat{v} = 0$ *so that* $\bar{X}$ *is the MRE.* □

The next result shows a least favorable property of the Gaussian distribution.

Theorem 7 *Let* $\mathcal{F}$ *be the class of all univariate distributions with density* f *w.r.t. Lebesgue measure and variance* $\sigma^2 = 1$*. Let* $X_1, \ldots, X_n$ *be i.i.d. distributed according to density in* $\mathcal{F}_f = \{f(x - \xi : \xi \in \mathbb{R})\}$ *with* $\xi = E(X_i)$*, with* f *fixed but arbitrary in* $\mathcal{F}$*. Assume* $L(\xi, d) = (d - \xi)^2$*, and let* δ *the MRE estimator of* ξ *(which exists since* L *is convex). Let*

$$r_n(F) = E_{F_0}(L(0, \delta(X))).$$

Then $r_n(F)$ *is largest when* F *is the Gaussian distribution.*

Proof. We showed that the MRE estimator in the normal case is $\tilde{\delta}(x) = \bar{X}$. The risk is

$$\begin{aligned}E_\xi((\tilde{\delta}(X) - \xi)^2) &= E_0((\bar{X})^2) \\ &= \frac{1}{n}.\end{aligned}$$

But this is the risk of $\tilde{\delta}(X)$ under f_0 for any $f \in \mathcal{F}$. Thus, since

$$\begin{aligned}\min_{\delta \in \Delta} R_{f_0}(0, \delta) &\leq R_{f_0}(0, \tilde{\delta}) \\ &= \min_{\delta \in \Delta} R_{f_0}(0, \delta),\end{aligned}$$

the theorem is proved. □

Recall that for squared error loss, the minimizer is

$$\hat{v}(y) = E_0(\delta_0(X)|Y = y),$$

and the MRE is

$$\hat{\delta}(x) = \delta_0(x) - E(\delta_0(X)|Y = y),$$

for an arbitrary fixed equivariant estimator.

Theorem 8 *(Pitman estimator) Assume $X_1, \ldots, X_n$ is a i.i.d. sample distributed according to location family, let f be (marginal) density and let $Y = (X_1 - X_n, \ldots, X_{n-1} - X_n)$, and assume $L(\xi, d) = (\xi - d)^2$. Then the MRE estimator $\hat{\delta}(x) = \delta_0(x) - E_0(\delta_0(X)|Y)$ is given by*

$$\hat{\delta}(x) = \frac{\int u f(x_1 - u, x_n - u)\, du}{\int f(x_1 - u, \ldots, x_n - u)\, du}$$

and called the Pitman estimator of ξ.

Proof. Let $\delta_0(x) = x_n$ and note that this is an equivariant estimator. Let $y_1 = x_1 - x_n, \ldots y_{n-1} = x_{n-1} - x_n, y_n = x_n$ which in matrix formulation is $Y = AX$ with matrix A having Jacobian $|A| = 1$. Then the joint density of $Y = (Y_1, \ldots, Y_n)$ is (by the change of variable formula)

$$p_Y(y_1, \ldots, y_n) = f(y_1 + y_n, \ldots, y_{n-1} + y_n, y_n),$$

and the conditional density of $\delta_0(X) = X_n = Y_n$ given $y = (y_1, \ldots, y_{n-1})$ is

$$\frac{f(y_1 + y_n, \ldots, y_{n-1} + y_n, y_n)}{\int f(y_1 + t, \ldots, y_{n-1} + t, t)\, dt}.$$

Therefore

$$\begin{aligned} E(\delta_0(X_n)|y) &= \frac{\int t f(y_1 + t, \ldots, y_{n-1} + t, t)\, dt}{\int f(y_1 + t, \ldots, y_{n-1} + t, t)\, dt} \\ &= \frac{\int t f(x_1 - x_n + t, \ldots, x_{n-1} - x_n + t, t)\, dt}{\int f(x_1 - x_n + t, \ldots, x_{n-1} - x_n + t, t)\, dt} \end{aligned}$$

which with change variable $u = x_n - t$ becomes

$$= x_n - \frac{\int u f(x_1 - u, \ldots, x_n - u)\, du}{\int f(x_1 - u, \ldots, x_n - u)\, du}.$$

Thus

$$\begin{aligned} \hat{\delta}(x) &= \delta_0(x) - E(\delta_0(X)|y) \\ &= \frac{\int u f(x_1 - u, \ldots, x_n - u)\, du}{\int f(x_1 - u, \ldots, x_n - u)\, du}, \end{aligned}$$

which ends the proof. $\square$

Example 24 *Assume $X_1, \ldots, X_n$ are i.i.d. $Un(\xi - b/2, \xi + b/2)$ with b assumed known and ξ unknown. The joint density is*

$$f(x_1 - \xi, \ldots, x_n - \xi) \;=\; \begin{cases} \frac{1}{b^n} \text{ if } \xi - \frac{b}{2} \leq x_{(1)} \leq x_1 \leq \ldots \leq x_n \leq x_{(n)} \leq \xi + \frac{b}{2}, \\ 0 \text{ otherwise.} \end{cases}$$

where $x_{(1)} \leq \ldots \leq x_{(n)}$ is the ordered sample. Assume we have quadratic loss function $\rho(u) = u^2$. Then the MRE is then given by the Pitman estimator as

$$\begin{aligned} \hat{\delta}(x) &= \frac{\int_{x_{(x)}-b/2}^{x_{(1)}+b/2} u b^{-n} du}{\int_{x_{(n)}-b/2}^{x_{(1)}+b/2} b^{-n} du} \\ &= \frac{\frac{u^2}{2}\Big|_{x_{(n)}-b/2}^{x_{(1)}+b/2}}{u\Big|_{x_{(n)}-b/2}^{x_{(1)}+b/2}} \\ &= \frac{1}{2}(x_{(1)} + x_{(n)}). \end{aligned}$$

□

5.3 Randomized estimators and equivariance

Randomized estimators $\tilde{\delta}(X)$ based on a sample X can be obtained using a deterministic rule δ

$$\tilde{\delta}(X) \;=\; \delta(X, W)$$

with W a r.v. that is independent of X and with a known distribution (i.e. a distribution that is not a function of the unknown parameter θ).

One can define invariance of location family distributions and loss functions as before

$$\begin{aligned} f(x'; \xi') &= f(x, \xi), \\ L(\xi', d') &= L(\xi, d), \end{aligned}$$

under transformations

$$\begin{aligned} x' &= x + a, \\ \xi' &= \xi + a, \\ d' &= d + a. \end{aligned}$$

One defines a randomized estimator to be equivariant if

$$\delta(X + a, W) \;=\; \delta(X, W) + a,$$

for all a. As before one can show that bias, variance and risk are all constant for such estimators.

It is easily seen that the set of equivariant estimators is given by

$$\{\delta(x,w) = \delta_0(x,w) + u(x,w) : u(x+a,w) = u(x,w), \quad \forall x, w, a\}$$

and with δ_0 a fixed equivariant estimator.

Again one can show that the condition $u(x+a,w) = u(x,w), \forall x, w, a$, holds if and anly if u is a function of $y = (x_1 - x_n, \ldots, x_{n-1} - x_n)$, so that $\delta(x,w)$ is equivariant if and only if

$$\delta(x,w) = \delta_0(x,w) - v(y,w).$$

Finally the MRE estimator is obtained by minimizing

$$E_0(\rho\,\{\delta_0(X,W) - v(Y,W)\}\,|Y = y, W = w).$$

But, one can start with any equivariant estimator, so we start with a *nonrandomized* equivariant estimator $\delta_0(X)$. Then

$$E_0\,(\rho(\delta_0(X) - v(Y,W))|Y = y, W = w) = E_0\,(\rho(\delta_0(X) - v(Y,W))|Y = y)\,,$$

since y is a function of x and X and W are independent. Then the minimizing function $\hat{v}$ will not depend on w, and therefore it will be nonrandomized. Thus the MRE estimator (if it exists) also when allowing for randomized estimators, will be nonrandomized.

Example 25 *Assume ρ is quadratic loss function. Then*

$$\hat{v} = \mathrm{argmin}_v E_0\,(\rho(\delta_0(X) - v(Y,W))|Y = y, W = w)$$

is given by

$$\begin{aligned} \hat{v}(y,w) &= E_0(\delta_0(X)|Y = y, W = w) \\ &= E_0(\delta(X)|Y = y) \end{aligned}$$

and is not a function of w. □

Thus, starting with a nonrandomized equivariant estimator $\delta_0(X)$, the MRE estimator $\delta_0(X) - \hat{v}(Y) =: \hat{\delta}(X)$ is nonrandomized.

5.3.1 Sufficiency and equivariance

Assume $\mathcal{P}$ is a location model family of distributions and that T is a sufficient statistic for ξ. Then any estimator $\delta(X)$ of ξ can be seen as a randomized estimator based on T, since there is always a randomized edstimator $\tilde{\delta}(T) = \delta(X)$ and thus

$$\delta(X) = \bar{\delta}(T,W)$$

for a deterministic rule $\bar{\delta}$, and with W a r.v. independent of T and with known distribution (so in particular not a function of ξ). Now assume that $T = (T_1, \ldots, T_r)$ and equivariant, so that

$$T(x+a) \quad = \quad T(x) + a.$$

Then one sees that the distribution of T is a location family (show this!), and we can view T as the original sample. Now since $\delta_0(X) = T(X)$ is equivariant and nonrandomized, one has that

$$\hat{v}(y, w) \quad = \quad \text{argmin}_{v:\mathbb{R}^{n-1}\to\mathbb{R}} E_0(\rho(\delta_0(T) - v(Y, W))|Y = y, W = w)$$

is a function only of y. Therefore the MRE estimator is given by

$$\hat{\delta}(T) \quad = \quad \delta_0(T) - \hat{v}(Y)$$

and is a function only of the sufficient statistic T.

□

This reasoning shows one connection between sufficiency and equivariance, namely that for a location family, if there is a sufficient statistic that is also equivariant, the the MRE estimator can be found to depend only on T.

Are MRE estimators unbiased?

Lemma 6 *Assume we have squared error loss. Then*

a) If $\delta(X)$ is an equivariant estimator with bias b, then $\delta(X) - b$ is equivariant, unbiased and has smaller risk that $\delta(X)$.

b) The (unique) MRE estimator is unbiased.

c) If a UMVU estimator exists and is equivariant then it is MRE.

Proof. a) Clearly $\delta(X) - b$ is equivariant and unbiased. Then since bias, risk, and variance does not depend on ξ

$$\begin{aligned} E_0((\delta(X) - b)^2) &= \text{Var}_0(\delta(X) - b) + \text{bias}(\delta(X) - b)^2 \\ &= \text{Var}_0(\delta(X)) \end{aligned}$$

which is (uniformly in ξ) smaller than

$$E_0((\delta(X))^2) \quad = \quad \text{Var}_0(\delta(X)) + \text{bias}(\delta(X))^2.$$

b) If $\delta(X)$ is the MRE and it is not unbiased, so if it would have a bias $b > 0$, then $\delta(X) - b$ would be unbiased, equivariant, with smaller risk which contradicts that $\delta(X)$ is the MRE.

c) An unbiased miminum risk estimator that is equivariant is by b) the MRE. □

Definition 13 *An estimator δ of $g(\theta)$ is called risk-unbiased if*

$$E_\theta L(\theta, \delta(X)) \quad \leq \quad E_\theta L(\theta', \delta(X)), \tag{1}$$

for all $\theta' \neq \theta$.

Example 26 *(Mean-unbiasedness) Assume we have squared error loss, and assume $\delta(X)$ is an estimator such that $E(\delta(X)^2) < \infty$. Then* (1) *is translated to*

$$E_\theta((\delta(X) - g(\theta))^2) \quad \leq \quad E_\theta((\delta(X) - g(\theta'))^2)$$

for all $\theta' \neq \theta$. Now assume that θ is fixed and let θ' vary and let us study the right hand side of (1)*. This is smallest, seen as a function of $g(\theta')$, for the value $g(\theta)$, by* (1)*. However, we know that the right hand side of* (1) *is minimized by $E_\theta(\delta(X))$. Since the loss function is strcitly convex the minimizing value is unique, and thus* (1) *is equivalent to*

$$g(\theta) \quad = \quad E_\theta(\delta(X)),$$

i.e. the "usual" definition of unbiasedness. □

The next result shows that MRE estimators are risk-unbiased.

Theorem 9 *Assume that δ is MRE for estimating ξ in a location invariant estimation problem. Then δ is risk-unbiased.*

Proof. The condition for risk-unbiasedness is

$$E_\xi \rho(\delta(X) - \xi') \quad \geq \quad E_\xi \rho(\delta(X) - \xi),$$

for all $\xi' \neq \xi$. Since the risk is constant for equivariant estimators, we can let $\xi = 0$ to obtain the condition

$$E_0 \rho(\delta(X) - a) \quad \geq \quad E_0 \rho(\delta(X)),$$

for all a. But $\delta(X)$ is the MRE estimator, so it has smaller risk than $\delta(X) - a$ for any a. Thus the condition for risk-unbiasedness is satisfied. □