Contents

This material comes from Chapter 6 of Berger [1985], Chapter 4 of Ferguson [1967], Chapter 3 Lehmann and Casella [1998] and Chapter 6 of Lehmann and Romano [2005].

1 Introduction

Suppose two scientists at different but nearby locations are tracking an object that emits a signal via a radio beacon. The first scientist will infer the location $\boldsymbol{\theta}_1$ of the object relative to himself via a measurement

$$\boldsymbol{X}_1 \sim N_p(\boldsymbol{\theta}_1, \sigma^2 \mathbf{I}),\ \boldsymbol{\theta}_1 \in \mathbb{R}^p, \sigma^2 > 0.$$

The second scientist would also like to infer the location $\boldsymbol{\theta}_2$ of the object relative to herself. She would like to estimate $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 + \boldsymbol{a}$, where $\boldsymbol{a}$ is the (known) displacement of scientist 2 from scientist 1. The inference problem from scientist 2's perspective can be expressed as

$$\begin{aligned}
&\boldsymbol{X}_2 = \boldsymbol{X}_1 + \boldsymbol{a},\ \boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 + \boldsymbol{a} \\
&\boldsymbol{X}_2 \sim N_p(\boldsymbol{\theta}_2, \sigma^2 \mathbf{I}),\ \boldsymbol{\theta}_2 \in \mathbb{R}^p, \sigma^2 > 0.
\end{aligned}$$

Note that the models for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are exactly the same. We say that the model for $\boldsymbol{X}_1$ is *invariant* under the transformation $\boldsymbol{X}_2 = g(\boldsymbol{X}_1) = \boldsymbol{X}_1 + \boldsymbol{a}$.

Now suppose each scientist wants to estimate their relative displacement from the object with squared error loss. Note that for any decision $\boldsymbol{d}_1$ about $\boldsymbol{\theta}_1$,

$$\begin{aligned}
L(\boldsymbol{\theta}_1, \boldsymbol{d}_1) = (\boldsymbol{\theta}_1 - \boldsymbol{d}_1)^2 &= ([\boldsymbol{\theta}_1 + \boldsymbol{a}] - [\boldsymbol{d}_1 + \boldsymbol{a}])^2 \\
&= (\boldsymbol{\theta}_2 - \boldsymbol{d}_2)^2 = L(\boldsymbol{\theta}_2, \boldsymbol{d}_2).
\end{aligned}$$

In other words, for every decision $\boldsymbol{d}_1$ one can make about $\boldsymbol{\theta}_1$, there is a corresponding decision $\boldsymbol{d}_2 = g(\boldsymbol{d}_1) = \boldsymbol{d}_1 + \boldsymbol{a}$ one can make about $\boldsymbol{\theta}_2$ that has the same loss profile as $\boldsymbol{d}_1$. As such, we say the loss is invariant. For this problem, we have

$$L(\boldsymbol{\theta}_1, \boldsymbol{d}_1) = L(g(\boldsymbol{\theta}_1), g(\boldsymbol{d}_1))\ \forall \boldsymbol{\theta} \in \mathbb{R}^p, \boldsymbol{d}_1 \in \mathbb{R}^p,$$

where $g(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{a}$.

Formal invariance:

The model and loss functions are exactly the same for the two scientists. Therefore if $\delta(\boldsymbol{X}_1)$ is the estimator of $\boldsymbol{\theta}_1$ for scientist 1, then $\delta(\boldsymbol{X}_2)$ is the estimator of $\boldsymbol{\theta}_2$ for scientist 2.

LC refer to this principle as *formal invariance*. The general idea is:

- estimating $\boldsymbol{\theta}_1$ from $\boldsymbol{X}_1$ is the same decision problem as estimating $\boldsymbol{\theta}_2$ from $\boldsymbol{X}_2$.
- Thus, our preferred estimator in each situation should be the same.

Suppose δ is a good estimator of $\boldsymbol{\theta}_1$ based on $\boldsymbol{X}_1$. Then we should have

$$\hat{\boldsymbol{\theta}}_2 = \delta(\boldsymbol{X}_2).$$

Now $\boldsymbol{X}_2 = \boldsymbol{X}_1 + \boldsymbol{a}$, so an estimate of $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 + \boldsymbol{a}$ can be obtained as

$$\widehat{\boldsymbol{\theta}_1 + \boldsymbol{a}} = \delta(\boldsymbol{X}_1 + \boldsymbol{a}).$$

Writing $\boldsymbol{\theta}_2 = g(\boldsymbol{\theta}_1)$ and $\boldsymbol{X}_1 + \boldsymbol{a} = g(\boldsymbol{X}_1)$ gives

$$\widehat{g(\boldsymbol{\theta}_1)} = \delta(g(\boldsymbol{X}_1)).$$

Formal invariance implies the estimate of the transformed parameter is the estimator at the transformed data.

Functional invariance:

Since $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_1 + \boldsymbol{a}$, if scientist 2 knew $\hat{\boldsymbol{\theta}}_1 = \delta(\boldsymbol{X}_1)$ she would estimate the location of the object as

$$\hat{\boldsymbol{\theta}}_2 = \hat{\boldsymbol{\theta}}_1 + \boldsymbol{a}$$
$$\widehat{g(\boldsymbol{\theta}_1)} = g(\delta(\boldsymbol{X}_1)).$$

The estimate of the transformed parameter is the transform of the parameter estimate.

The principle of invariance:

The principle of invariance is to require a functionally invariant estimator in a formally invariant problem. For this particular problem, this means we should use an estimator that satisfies

$$\delta(g(\boldsymbol{x})) = g(\delta(\boldsymbol{x}))\ \forall \boldsymbol{x} \in \mathbb{R}^p, \text{ that is,}$$
$$\delta(\boldsymbol{x} + \boldsymbol{a}) = \delta(\boldsymbol{x}) + \boldsymbol{a}\ \forall \boldsymbol{x} \in \mathbb{R}^p.$$

Such an estimator is called *equivariant with respect to the transformation g*.

Best equivariant estimators

It seems logical that we would want this to hold regardless of $\boldsymbol{a} \in \mathbb{R}^p$. If so, then we require

$$\delta(\boldsymbol{x} + \boldsymbol{a}) = \delta(\boldsymbol{x}) + \boldsymbol{a} \ \forall \boldsymbol{x} \in \mathbb{R}^p, \boldsymbol{a} \in \mathbb{R}^p.$$

An estimator satisfying this condition is equivariant with respect to the (group of) transformations $\mathcal{G} = \{g : \boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{a}, \boldsymbol{a} \in \mathbb{R}^p\}$. For this problem, the set equivariant estimators is easy to identify: Setting $\boldsymbol{a} = -\boldsymbol{x}$, we have

$$\begin{aligned}
\delta(\boldsymbol{x} - \boldsymbol{x}) &= \delta(\boldsymbol{x}) - \boldsymbol{x} \\
\delta(\boldsymbol{x}) &= \boldsymbol{x} + \delta(\boldsymbol{0}) \\
\delta(\boldsymbol{x}) &= \boldsymbol{x} + \boldsymbol{c},
\end{aligned}$$

i.e., $\delta(\boldsymbol{x})$ is equivariant with respect to $\mathcal{G}$ if and only if $\delta(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{c}$ for some $\boldsymbol{c} \in \mathbb{R}^p$. Of course, we would like to find the best equivariant estimator, i.e. the one that minimizes the risk. Since

$$\begin{aligned}
R(\boldsymbol{\theta}, \boldsymbol{x} + \boldsymbol{c}) &= \mathrm{E}[\sum_{j=1}^{p} (X_j + c_j - \theta_j)^2] \\
&= \sum_{j=1}^{p} \mathrm{Var}[X_j + c_j] + \mathrm{Bias}^2[X_j + c_j] \\
&= p\sigma^2 + \boldsymbol{c}^T \boldsymbol{c},
\end{aligned}$$

the best equivariant estimator is $\boldsymbol{x}$, obtained by setting $\boldsymbol{c} = \boldsymbol{0}$. Thus $\delta(\boldsymbol{x}) = \boldsymbol{x}$ is the UMRE (uniformly minimum risk equivariant) estimator.

More generally, restricting attention to equivariant estimators can reduce the class of candidate estimators considerably, often enough so that a single optimal estimator may be identified.

Does equivariance make sense? Often yes, especially if we lack information about the unknown parameter. However, if we do have some information about the parameter, then the problem may not be invariant and an equivariant estimator may not make sense.

Example:

Consider the model $\boldsymbol{X} \sim N_p(\boldsymbol{\theta}, \sigma^2\mathbf{I}) : \sigma^2 > 0,\ \boldsymbol{\theta} \in A \subset \mathbb{R}^p$.
This model is *not* invariant under $g : \boldsymbol{x} \to \boldsymbol{x}+\boldsymbol{a}$. Letting $\boldsymbol{X}' = g(\boldsymbol{X})$ and $\boldsymbol{\theta}' = g(\boldsymbol{\theta}')$,

- The models for $\boldsymbol{X}$ and $\boldsymbol{X}'$ are not the same,
- Our estimator of $\boldsymbol{\theta}$ should be in A w.p. 1.
- Our estimator of $\boldsymbol{\theta}'$ should be in $A + \boldsymbol{a}$ w.p. 1.

Thus we lack formal invariance. We may still want functional invariance, that is

$$\delta'(\boldsymbol{x}') = \hat{\boldsymbol{\theta}}' = \hat{\boldsymbol{\theta}} + \boldsymbol{a} = \delta(\boldsymbol{x}) + \boldsymbol{a},$$

but without formal invariance we wouldn't use the same estimator in the two situations. In particular, we wouldn't necessarily want

$$\delta(\boldsymbol{x} + \boldsymbol{a}) = \delta(\boldsymbol{x}) + \boldsymbol{a},$$

especially if $A \cap \{A + \boldsymbol{a}\} = \emptyset$.

2 Invariant estimation problems

2.1 Invariance under a transformation

Let $\mathcal{P}$ be a statistical model for $\{\mathcal{X}, \sigma(\mathcal{X})\}$.
Let g be a bijection on $\mathcal{X}$ (one-one and onto).

Definition 1 (model invariance under a transformation)**.** *$\mathcal{P}$ is invariant to g if*

$$\text{when } X \sim P \in \mathcal{P}, \text{ then } gX = X' \sim P' \in \mathcal{P}$$

and

$$\forall P' \in \mathcal{P}, \ \exists P \in \mathcal{P} : X \sim P \to gX \sim P'.$$

The latter condition ensures that g doesn't "reduce" the model.
Informally, $\mathcal{P}$ is invariant to g if "$g\mathcal{P} = \mathcal{P}$" (but this notation isn't quite right).

Example:
Let $g(x) = 2x$, and

$$\mathcal{P}_A = \{\text{dnorm}(x, \mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$$
$$\mathcal{P}_B = \{\text{dnorm}(x, \mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 1\}$$

Then $\mathcal{P}_A$ is invariant to g but $\mathcal{P}_B$ is not. The transformation "reduces" $\mathcal{P}_B$. An estimator appropriate for $\mathcal{P}_B$ may not be appropriate for "$g\mathcal{P}_B$". An estimator for the former model should allow $\hat{\sigma^2} \in [1, 4]$, an estimator for the latter model should not.

Induced transformation
Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a (parametric) model that is invariant to g.
When $X \sim P_\theta$, $\theta \in \Theta$, then $gX \sim P_{\theta'}$ for some $\theta' \in \Theta$.

Define $\bar{g}$ to be this transformation of the parameter space:

$$\bar{g}\theta = \theta' \ : X \sim P_\theta \to gX \sim P_{\theta'}.$$

The function $\bar{g}$ is only well defined if the parametrization is identifiable, so that $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$. We will assume we are working with an identifiable parametrization in what follows. In particular, identifiability makes $\bar{g}$ a bijection:

Lemma 1. *If $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is invariant under a bijection g and the parametrization is identifiable, then $\bar{g}$ is a bijection.*

Proof. That $\bar{g}\Theta = \Theta$ (g is onto) follows from the second condition in the definition of invariance: $\forall \theta_1 \in \Theta \; \exists \theta_0 \in \Theta : \bar{g}\theta_0 = \theta_1$. That $\bar{g}$ is 1-1 follows from the identifiability: Suppose $\bar{g}\theta_1 = \bar{g}\theta_2 = \theta_0$. Then $\forall A \in \sigma(\mathcal{X})$

$$\begin{aligned} \Pr(gX \in A|\theta_1) &= \Pr(gX \in A|\theta_2) \\ \Pr(X \in g^{-1}A|\theta_1) &= \Pr(X \in g^{-1}A|\theta_2). \end{aligned}$$

This shows that $\Pr(X \in B|\theta_1) = \Pr(X \in B|\theta_2)$ for all sets B of the form $\{B = g^{-1}A : A \in \sigma(\mathcal{X})\}$.
This includes all of $\sigma(\mathcal{X})$: Pick any $B \in \sigma(\mathcal{X})$ and let $A = gB$, so $g^{-1}A = g^{-1}gB = B$ as g is a bijection.
Now since $\Pr(X \in B|\theta_1) = \Pr(X \in B|\theta_2) \; \forall B \in \sigma(X)$ and the parametrization is identifiable, we must have $\theta_1 = \theta_2$. Thus $\bar{g}\theta_1 = \bar{g}\theta_2$ iff $\theta_1 = \theta_2$. $\square$

We now have proper notation to describe invariance of $\mathcal{P}$ under a transformation g of X:

$$\mathcal{P} \text{ is invariant under } g \text{ if } \bar{g}\Theta = \Theta.$$

The following set of identities will be useful:

Lemma 2. *If $\mathcal{P}$ is invariant to g, then*

$$\begin{aligned} \Pr(gX \in A|\theta) &= \Pr(X \in A|\bar{g}\theta) \\ \Pr(X \in gA|\bar{g}\theta) &= \Pr(X \in A|\theta) \end{aligned}$$

Note that each statement follows from the definition, and each implies the other. For example, to show that the first implies the second, note that

$$\begin{aligned} \Pr(X \in A|\theta) &= \Pr(gX \in gA|\theta) \\ &= \Pr(X \in gA|\bar{g}\theta). \end{aligned}$$

The first equality follows from g being a bijection, the second from the first identity in the Lemma.
<u>Example:</u>

$\mathcal{P} = \{\text{dnorm}(x, \mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

$g(X) = a + bX$

$\bar{g}(\mu, \sigma^2) = (a + b\mu, b^2\sigma^2)$

$A = (-\infty, c]$

Then we have

$$\begin{aligned}\Pr(gX \in A|\theta) &= \Pr(a + bX \leq c|(\mu, \sigma^2)) \\ &= \Pr(X \leq c|(a + b\mu, b^2\sigma^2)) = \Pr(X \in A|\bar{g}(\theta)).\end{aligned}$$

Going the other way,

$$\begin{aligned}\Pr(X \in gA|\bar{g}\theta) &= \Pr(X \leq a + bc|a + b\mu, b^2\sigma^2) \\ &= \Pr([X - a]/b \leq c|a + b\mu, b^2\sigma^2) \\ &= \Pr(X \leq c|\mu, \sigma^2) = \Pr(X \in A|\theta).\end{aligned}$$

<u>Invariant loss:</u>
If the model is invariant under g it is natural to require the loss be invariant in some sense as well.

- $\{L(\theta, d) : \theta \in \Theta, d \in D\}$ gives, for each θ, the pairings of losses to decisions about θ.
- $\{L(\bar{g}\theta, d) : \theta \in \Theta, d \in D\}$ gives, for each θ, the pairings of losses to decisions about $\bar{g}\theta$.

If estimation of θ and $\bar{g}\theta$ are the same problem, then the set of possible losses at θ should correspond to the possible losses under $\bar{g}\theta$.

Definition 2 (loss invariant under a transformation)**.** *Let $\mathcal{P}$ be invariant under g, so $\bar{g}\Theta = \Theta$. A loss function $L(\theta, d) : \Theta \times D \to \mathbb{R}^+$ is invariant if*

$$\forall d \in D \;\; \textit{there exists a unique } d' \in D : L(\theta, d) = L(\bar{g}\theta, d') \;\forall \theta \in \Theta.$$

The decision d' that corresponds to d is referred to as $\tilde{g}d$, which is a bijection on D. Invariance of the loss under $\tilde{g}$ then means that

$$L(\theta, d) = L(\bar{g}\theta, \tilde{g}d) \ \forall \ \theta \in \Theta, d \in D.$$

Nonexample:
$\mathcal{P} = \{\text{dnorm}(x, \mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$
$g(X) = a + bX$
$\bar{g}(\mu, \sigma^2) = (a + b\mu, b^2\sigma^2)$
If loss is squared error loss, then $L(\theta, d) = (\mu - d)^2$ and $L(\bar{g}\theta, d') = (a + b\mu - d')^2$. Invariance of loss means for each d, there is a d' such that these are equal for all θ. Is this possible? First try setting $\mu - d = a + b\mu - d'$:

$$\begin{aligned} \mu - d &= a + b\mu - d' \\ d' &= d + a + \mu(b - 1). \end{aligned}$$

Clearly this will not work if $b \neq 1$ - there is no d' that has the same loss profile as d.

Example:
Consider the same example but with standardized loss

$$L(\theta, d) = (\mu - d)^2/\sigma^2.$$

Equating losses $L(\theta, d)$ to $L(\bar{g}\theta, d')$, one solution is given by

$$\begin{aligned} (\mu - d)/\sigma &= (a + b\mu - d')/[b\sigma] \\ \mu - d &= a/b + \mu - d'/b \\ d' &= a + bd. \end{aligned}$$

The loss of decision d' under $\bar{g}\theta$ is the same as that of decision d under θ, for all θ. The loss function is invariant, and the induced transformation on the decision space is

$$\tilde{g}d = a + bd.$$

We say that a decision problem is invariant under g if the model and loss are invariant under the induced transformations $\bar{g}$ and $\tilde{g}$:

Definition 3 (invariant decision problem)**.** *A decision problem* (Θ, D, L) *is invariant under* g *if*

- *the parameter space is invariant under the induced transformation* $\bar{g}$
- *the loss is invariant under the induced transformation* $\tilde{g}$

2.2 Invariance under a group:

Typically if a problem is invariant under a particular transformation g, it is also invariant under a a class of related transformations. This class can always taken to be a group, meaning that if (Θ, D, L) is invariant under a class $\mathcal{C}$, it is invariant under a group $\mathcal{G}$ that is generated by $\mathcal{C}$.

Definition 4 (group)**.** *A collection* $\mathcal{G}$ *of one-to-one transformation of* $\mathcal{X}$ *is a group if*

1. $\forall g_1, g_2 \in \mathcal{G}$, $g_1 g_2 \in \mathcal{G}$ *(closure under composition),*
2. $\forall g \in \mathcal{G}$, $g^{-1} \in \mathcal{G}$ *(closure under inversion),*
3. *the function* $gx = x$ *is in* $\mathcal{G}$*.*

Example (linear transformations):
Let $\mathcal{X} = \mathbb{R}$ and let

$$\mathcal{G} = \{g : g(x) = a + bx, a \in \mathbb{R}, b \in \mathbb{R} \setminus \{0\}\}.$$

Check to see if this is a group:

- Each g is 1-1 (as we don't allow $b = 0$).

- Let $g_j(x) = a_j + b_j x$, for $j \in \{1, 2\}$. Then

$$g_1(g_2(x)) = a_1 + b_1(a_2 + b_2 x) = (a_1 + a_2 b_1) + (b_1 b_2)x \in \mathcal{G}.$$

- Let $g(x) = a + bx$. Then

$$g^{-1}(x) = -a/b + x/b \in \mathcal{G}.$$

So yes, this is a group.

A decision problem is invariant under a group if it is invariant under each function in the group.

Example (normal mean estimation):
$\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$.
$L(\theta, d) = (\mu - d)^2/\sigma$ for $d \in D = \mathbb{R}$
$\mathcal{G} = \{g : g(x) = a + bx, a \in \mathbb{R}, b \in \mathbb{R} \setminus \{0\}\}$.
For a single $g(x) = a + bx \in \mathcal{G}$ the induced transformations on Θ and D are

- $\bar{g}(\mu, \sigma^2) = (a + b\mu, b^2\sigma^2)$;
- $\tilde{g}d = a + bd$.

We have also shown invariance for each $g \in \mathcal{G}$:

- $\bar{g}\Theta = \Theta$
- $L(\theta, d) = L(\bar{g}\theta, \tilde{g}d)\ \forall \theta \in \Theta,\ d \in D$.

Thus we say the decision problem (Θ, D, L) is invariant under $\mathcal{G}$.
Note that things are essentially unchanged if we consider the n-sample problem:

$$\mathcal{P} = \{N_n(\mu\mathbf{1}, \sigma^2\mathbf{I}) : \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}.$$

Induced groups:
Note that a group $\mathcal{G}$ under which a problem is invariant induces two other collections of functions:

- $\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}$, transformations on the parameter space;
- $\tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\}$, transformations on the decision space.

As you might suspect, these collections turn out to be groups:

Lemma 3. *If $\{P_\theta : \theta \in \Theta\}$ is invariant under a group $\mathcal{G}$, then*

- *$\bar{\mathcal{G}} = \{\bar{g} : g \in \mathcal{G}\}$ is a group of transformations of Θ onto itself;*
- *$\tilde{\mathcal{G}} = \{\tilde{g} : g \in \mathcal{G}\}$ is a group of transformations of D onto itself.*

The functions in $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are automatically 1-1 and onto.
If you are interested in such things, note that $\bar{\mathcal{G}}$ is a homomorphic image of $\mathcal{G}$, and $\tilde{\mathcal{G}}$ is a homomorphic image of $\mathcal{G}$ and $\bar{\mathcal{G}}$.

Example (scale group):
$\mathcal{P} = \{p(x|\theta) = e^{-x/\theta}/\theta : x > 0, \theta \in \Theta = \mathbb{R}^+\}$
$\mathcal{G} = \{g_c : g_c(x) = cx, c \in \mathbb{R}^+\}$
$L(\theta, d) = (1 - d/\theta)^2$

- Invariance:
  - If $X \sim P_\theta, \theta \in \Theta$ and $X' = g_c(X)$, then $X' \sim P_{\theta'}$, $\theta' = c\theta \in \Theta$.
  - If $X' \sim P_{\theta'}$ then $X' \stackrel{d}{=} g_c X$ for $X \sim P_\theta$, $\theta = \theta'/c$, which is in Θ for all θ', c.
- Induced group on Θ: $\bar{\mathcal{G}} = \{\bar{g}_c\theta = c\theta : c \in \mathbb{R}^+\}$.
- Induced group on D: Solve $L(\theta, d) = L(\bar{g}_c\theta, d')$ for d'

$$(1 - d/\theta)^2 = (1 - d'/\bar{g}_c\theta)^2$$
$$d/\theta = d'/c\theta$$
$$d' = cd.$$

Thus for every d, there is a d' such that $L(\theta, d) = L(\bar{g}\theta, d')$, so the loss is invariant. Defining $d' = \tilde{g}_c d$, we have $\tilde{\mathcal{G}} = \{\tilde{g}_c : \tilde{g}_c d = cd, c \in \mathbb{R}^+\}$.

Note that $\mathcal{G}$, $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are all the same group (the multiplicative group, or scale group).

Example (location group):
$\mathcal{P} = \{\text{dnorm}(x|\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}$
$\mathcal{G} = \{g_c : g_c(x) = x + c, c \in \mathbb{R}\}$
$L(\theta, d) = f(|\mu - d|)$

- Invariance:
  - If $X \sim P_\theta, \theta \in \Theta$ and $X' = g_c(X)$, then $X' \sim P_{\theta'}$, $\theta' = (\mu', \sigma^{2'}) = (\mu + c, \sigma^2) \in \Theta$.
  - If $X' \sim P_{\theta'}$ then $X' \stackrel{d}{=} g_c X$ for $X \sim P_\theta$, $\theta = \theta' - (c, 0)$, which is in Θ for all θ', c.
- Induced group on Θ: $\bar{\mathcal{G}} = \{\bar{g}_c\theta = \theta + (c, 0) : c \in \mathbb{R}\}$.
- Induced group on D: Solve $L(\theta, d) = L(\bar{g}_c\theta, d')$ for d'

$$\begin{aligned} f(|\mu - d|) &= f(|\mu' - d'|) \\ \mu - d &= \mu + c - d' \\ d' &= d + c. \end{aligned}$$

  Thus for every d, there is a d' such that $L(\theta, d) = L(\bar{g}\theta, d')$, so the loss is invariant. Defining $d' = \tilde{g}_c d$, we have $\tilde{\mathcal{G}} = \{\tilde{g}_c : \tilde{g}_c d = d + c, c \in \mathbb{R}^+\}$.

Note that $\mathcal{G}$ and $\tilde{\mathcal{G}}$ are both the location (additive) group.
The group $\bar{\mathcal{G}}$ is isomorphic to these groups.

Example (covariance estimation):
Let $\mathbf{X}$ be the $n \times p$ matrix with rows $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim$ i.i.d. $N_p(\boldsymbol{\mu}, \Sigma)$.
We say $\mathbf{X}$ has a matrix normal distribution, $\mathbf{X} \sim N_{n \times p}(\mathbf{1}\boldsymbol{\mu}^T, \mathbf{I}, \Sigma)$
$\mathcal{P}$ = distributions of $\mathbf{X} \sim N_{n\times p}(\mathbf{1}\boldsymbol{\mu}^T, \mathbf{I}, \Sigma)$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\Sigma \in \mathcal{S}_p^+$.
$\mathcal{G} = \{g : g(\mathbf{x}) = \mathbf{1}\boldsymbol{a}^T + \mathbf{x}\mathbf{B}^T, \boldsymbol{a} \in \mathbb{R}^p, \mathbf{B} \in \mathbb{R}^{p\times p}, \text{invertible}\}$.
$L(\theta, \mathbf{D}) = \text{tr}(\mathbf{D}\Sigma^{-1}) - \log|\mathbf{D}\Sigma^{-1}| - p$ (Stein's loss)

- Invariance:
  - If $\mathbf{X} \sim P_\theta$, $\theta \in \Theta$ and $\mathbf{X}' = g(X)$, then
$$\mathbf{X}' \sim P_{\theta'},\ \theta' = (\boldsymbol{\mu}', \Sigma') = (\boldsymbol{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}^T) \in \Theta.$$
  - If $\mathbf{X}' \sim P_{\theta'}$ then $\mathbf{X}' \stackrel{d}{=} gX$ for $X \sim P_\theta$, where
$$\theta = (\boldsymbol{\mu}, \Sigma) = (\mathbf{B}^{-1}(\boldsymbol{\mu}' - \boldsymbol{a}), \mathbf{B}^{-1}\Sigma'\mathbf{B}^{-1}),$$
    which is in Θ for all θ', $\boldsymbol{a}$, $\mathbf{B}$.
- Induced group on Θ: $\bar{\mathcal{G}} = \{\bar{g}\theta = g(\boldsymbol{\mu}, \Sigma) = (\boldsymbol{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}^T\}, \boldsymbol{a} = \mathbb{R}^p, \mathbf{B} \text{ invertible}\}$.
- Induced group on D: Solve $L(\theta, \mathbf{D}) = L(\bar{g}\theta, \mathbf{D}')$ for $\mathbf{D}'$
$$\begin{aligned}\text{tr}(\mathbf{D}\Sigma^{-1}) - \log|\mathbf{D}\Sigma^{-1}| &= \text{tr}(\mathbf{D}'\mathbf{B}^{-1T}\Sigma^{-1}\mathbf{B}^{-1}) - \log|\mathbf{D}'\mathbf{B}^{-1T}\Sigma^{-1}\mathbf{B}^{-1}| \\ &= \text{tr}(\mathbf{B}^{-1}\mathbf{D}'\mathbf{B}^{-1T}\Sigma^{-1}) - \log|\mathbf{B}^{-1}\mathbf{D}'\mathbf{B}^{-1T}\Sigma^{-1}|,\end{aligned}$$
  and so equality for all $\theta \in \Theta$ can be achieved by setting $\mathbf{D}' = \mathbf{B}\mathbf{D}\mathbf{B}^T$. Thus for every $\mathbf{D}$, there is a $\mathbf{D}'$ such that $L(\theta, \mathbf{D}) = L(\bar{g}\theta, \mathbf{D}')$, so the loss is invariant. Defining $\mathbf{D}' = \tilde{g}\mathbf{D}$, we have $\tilde{\mathcal{G}} = \{\tilde{g} : \tilde{g}\mathbf{D} = \mathbf{B}\mathbf{D}\mathbf{B}^T : \mathbf{B} \text{ invertible}\}$.

Here, the groups $\mathcal{G}$, $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ are different.
$\mathcal{G}$ and $\bar{\mathcal{G}}$ are isomorphic to each other, but not to $\tilde{\mathcal{G}}$.

3 Invariant decision rules

Suppose $\{\Theta, D, L\}$ is invariant under a group $\mathcal{G}$.

X Decision problem:

$X \sim P_\theta : \theta \in \Theta$

Estimate θ with loss L

X' Decision problem:

$X' = gX$, $\theta' = \bar{g}\theta$.

$X' \sim P_{\theta'} : \theta' \in \Theta$

Estimate θ' with loss L

Invariance principle:

$X' = gX$, $\theta' = \bar{g}\theta$, $L(\theta, d) = L(g\theta, \tilde{g}d)$

- Formal invariance: The models and losses are the same. We would use the same estimator in either situation.

$$\hat{\theta'} = \delta(x') = \delta(gx)$$

- Functional invariance:
  - When $X = x$ we make decision $\delta(x)$ and incur loss $L(\theta, \delta(x))$.
  - By invariance of loss, estimating θ by $\delta(x)$ incurs the same loss as estimating $\theta' = \bar{g}\theta$ by $\tilde{g}\delta(x)$:

$$L(\theta, \delta(x)) = L(\bar{g}\theta, \tilde{g}\delta(x)).$$

This suggests that if we are happy with estimating θ by $\delta(x)$, we should be happy estimating $\theta' = \bar{g}\theta$ by $\tilde{g}\delta(x)$.

$$\hat{\theta'} = \tilde{g}\delta(x).$$

Combining the two requirements gives

$$\delta(gx) = \tilde{g}\delta(x).$$

Definition 5. *For decision problem invariant under a group $\mathcal{G}$, an estimator is equivariant if*

$$\delta(gx) = \tilde{g}\delta(x) \ \forall g \in \mathcal{G}.$$

If we have decided that we want to use an equivariant estimator, the tasks are to

1. characterize the equivariant estimators;

2. identify the equivariance estimator with minimum risk.

That this second task is well defined is suggested by the following theorem:

Theorem 1. *The risk of an equivariant decision rule satisfies*

$$R(\theta, \delta) = R(\bar{g}\theta, \delta) \ \forall \theta \in \Theta, \bar{g} \in \bar{\mathcal{G}}.$$

Proof.

$$\begin{aligned} R(\theta, \delta) &= \mathrm{E}[L(\theta, \delta(X))|\theta] \\ &= \mathrm{E}[L(\bar{g}\theta, \tilde{g}\delta(X))|\theta] \quad \text{(invariance of loss)} \\ &= \mathrm{E}[L(\bar{g}\theta, \delta(gX))|\theta] \quad \text{(equivariance of } \delta) \\ &= \mathrm{E}[L(\bar{g}\theta, \delta(X))|\bar{g}\theta] \quad \text{(recall } \Pr(gX \in A|\theta) = \Pr(X \in A|g\theta)) \\ &= R(\bar{g}\theta, \delta). \end{aligned}$$

□

Interpretation:

If $\theta' = \bar{g}\theta$, then θ and θ' should be equally difficult to estimate (i.e. are equally risky).

We can define equivalence classes of equally risky θ-values:

Definition 6. *Two points $\theta_0, \theta_1 \in \Theta$ are* equivalent *if $\theta_1 = \bar{g}\theta_0$ for some $\bar{g} \in \bar{\mathcal{G}}$. The* orbit *$\Theta(\theta_0)$ of $\theta_0 \in \Theta$ is the set of equivalent points:*

$$\Theta(\theta_0) = \{\bar{g}(\theta_0) : \bar{g} \in \bar{\mathcal{G}}\}.$$

The above theorem should then be interpreted as saying that the risk function of an equivariant estimator is constant on orbits of Θ. In most of our applications, there is only one orbit of Θ, i.e. the class $\bar{\mathcal{G}}$ is rich enough so that we can go from any point in Θ to another via some $\bar{g} \in \mathcal{G}$. In such cases, $\bar{\mathcal{G}}$ is said to be *transitive* over Θ.

Corollary 1. *If $\bar{\mathcal{G}}$ is transitive, then the risk function of any equivariant estimator is constant over the parameter space.*

In such cases, the risk function of each equivariant estimator reduces to a single number, and finding the (uniformly) minimum risk equivariant estimator amounts to minimizing this single number.

Exercise: Review these ideas in the context of the vector location problem at the beginning of these notes.

4 Simple examples

Example (location family):
$X \sim P_\theta \in \mathcal{P} = \{p_0(x - \theta) : \theta \in \mathbb{R}\}$
$L(\theta, d) = (\theta - d)^2$
$\mathcal{G} = \{g : x \to x + a, a \in \mathbb{R}\}$

This decision problem is invariant under $\mathcal{G}$, with

$$g_a(x) = x + a,\ \bar{g}_a(\theta) = \theta + a,\ \tilde{g}_a(d) = d + a.$$

Also note that $\bar{\mathcal{G}}$ is transitive over Θ, so any equivariant estimator must have constant risk. Let's check this: Any equivariant estimator must satisfy

$$\delta(gx) = \tilde{g}\delta(X)$$
$$\delta(x + a) = \delta(x) + a.$$

This must hold for all $x, a \in \mathbb{R}$. To characterize such estimators as a function of x, pick $a = -x$ to show that

$$\delta(x) = x + \delta(0) \equiv x + c.$$

Finding the best equivariant estimator is then simply finding the value of c that minimizes the risk. What are the risk functions of such estimators?

$$\begin{aligned} R(\theta, \delta) &= R(0, \delta) \\ &= \mathrm{E}[L(0, X + c)|0] \\ &= \int (x + c)^2 p_0(x)\ dx = \mathrm{E}[X^2|0] + 2c\mathrm{E}[X|0] + c^2 \end{aligned}$$

which is minimized by setting $c = -\mathrm{E}[X|0]$. The minimum risk equivariant estimator is then

$$\delta(x) = x - \mathrm{E}[X|0].$$

If p_0 has a zero mean, then we have $\delta(x) = x$.

<u>Exercise:</u> Show that if $L(\theta, d) = |\theta - d|$, then the problem is still invariant, and the best estimator is $\delta(x) = x - c$, where c is the median of p_0.

There is an alternative way to interpret this result: As shown above, the best equivariant estimator is obtained by minimizing $\mathrm{E}[(X + c)^2|0]$ in c. This expectation can be rewritten as

$$\begin{aligned} \mathrm{E}[(X + c)^2|0] &= \int (x' + c)^2 p_0(x')\ dx' \\ &= \int (x' + (d_x - x))^2 p_0(x')\ dx' && \text{(decision under } x \text{ is } d_x = x + c) \\ &= \int ((x - \theta) + (d_x - x))^2 p_0(x - \theta)\ d\theta && \text{(change of variables: } \theta = x - x'\ ) \\ &= \int (\theta - d_x)^2 p_0(x - \theta)\ d\theta \\ &= \int (\theta - d_x)^2 \pi(\theta|x)\ d\theta = R(\pi, d_x|x), \end{aligned}$$

where $\pi(\theta|x)$ is the generalized Bayes posterior distribution under the uniform "prior" distribution on $\mathbb{R}$:

$$\begin{aligned}\pi(\theta|x) &= \frac{\pi(\theta)p(x|\theta)}{\int \pi(\theta)p(x|\theta)\ d\theta} \\ &= \frac{p_0(x-\theta)}{\int p_0(x-\theta)\ d\theta} \\ &= \frac{p_0(x-\theta)}{\int p_0(x-\theta)\ dx} = p_0(x-\theta).\end{aligned}$$

The generalized Bayes rule is the minimizer of this generalized Bayes risk. Our previous calculation has shown that this generalized Bayes rule is $\delta_\pi(x) = d_x = x+c$, where $c = -\mathrm{E}[X|0]$. Re-deriving this from the Bayesian perspective, the generalized Bayes rule under squared error loss is the posterior mean:

$$\begin{aligned}\delta_\pi(x) = \mathrm{E}[\theta|x] &= \int \theta\pi(\theta|x)\ d\theta \\ &= \int \theta p_0(x-\theta)\ d\theta \\ &= \int (x-x')p_0(x')\ dx' \qquad \text{(change of variables: } x' = x-\theta \text{)} \\ &= x - \int x'p_0(x')\ dx' = x - \mathrm{E}[X|0].\end{aligned}$$

The best equivariant estimator (under both squared error and absolute loss) is therefore equal to the generalized Bayes estimator under the improper prior $\pi(\theta) \propto 1$. It is no coincidence that this prior has invariance properties from a Bayesian perspective. We will see further examples of this in the upcoming material.

Note: To make the above change of variables, you can do it the "calculus way" via substitution (be sure to change the limits of integration), or the "probability way". For this latter approach, note that we want to compute the expectation of θ with respect to the density $p_0(x-\theta)$. Letting $X' = x-\theta$, computing the expectation of θ is the same as computing the expectation of $x - X'$. The pdf for x' is obtained from that of θ via the usual change of variables formula: $p_{x'}(x') = p_\theta(\theta(x'))|\frac{d\theta}{dx'}| = p_0(x-[x-x']) \times 1 = p_0(x')$.

Exercise: Obtain the UMRE estimator for the vector location problem $\boldsymbol{X} \sim p_0(\boldsymbol{x}-\boldsymbol{\theta})$ under the group $\mathcal{G} = \{g_a : \boldsymbol{x} \to \boldsymbol{x} + \boldsymbol{a}, \boldsymbol{a} \in \mathbb{R}^p\}$.

Example (covariance estimation):
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim$ i.i.d. $N_p(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} \in \mathbb{R}, \Sigma \in \mathcal{S}_p^+$.
Then $\mathbf{S} = \sum(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T \sim \text{Wishart}(n-1, \Sigma), \Sigma \in \mathcal{S}_p^+$.

Consider estimating Σ based on $\mathbf{S}$, under Stein's loss $L(\theta, \mathbf{D}) = \text{tr}(\mathbf{D}\Sigma^{-1}) - \log|\mathbf{D}\Sigma^{-1}| - p$. It is straightforward to show that the estimation problem is invariant under transformations in $\mathcal{G} = \{g : \mathbf{S} \to \mathbf{B}\mathbf{S}\mathbf{B}^T : \mathbf{B} \in \mathbb{R}^{p\times p}, \text{nonsingular}\}$. These transformations induce the following groups on $\Theta = D = \mathcal{S}_p^+$:

- $\bar{\mathcal{G}} = \{\bar{g} : \Sigma \to \mathbf{B}\Sigma\mathbf{B}^T, \mathbf{B} \in \mathbb{R}^{p\times p}, \text{nonsingular}\}$;
- $\tilde{\mathcal{G}} = \{\tilde{g} : \mathbf{D} \to \mathbf{B}\mathbf{D}\mathbf{B}^T, \mathbf{B} \in \mathbb{R}^{p\times p}, \text{nonsingular}\}$;

Note that all groups are the same, and operate on the same space.

An equivariant estimator must satisfy

$$\delta(g\mathbf{S}) = \tilde{g}\delta(\mathbf{S})$$
$$\delta(\mathbf{B}\mathbf{S}\mathbf{B}^T) = \mathbf{B}\delta(\mathbf{S})\mathbf{B}^T \ \forall \mathbf{S}, \mathbf{B}.$$

To characterize the equivariant estimators, choose $\mathbf{B} = \mathbf{S}^{-1/2}$, where $\mathbf{S}^{1/2}\mathbf{S}^{1/2} = \mathbf{S}$. This gives

$$\delta(\mathbf{I}) = \mathbf{S}^{-1/2}\delta(\mathbf{S})\mathbf{S}^{-1/2}$$
$$\delta(\mathbf{S}) = \mathbf{S}^{1/2}\delta(\mathbf{I})\mathbf{S}^{1/2}.$$

We still need to characterize the possibilities for $\delta(\mathbf{I})$. Recall an orthogonal $p \times p$ matrix $\mathbf{U}$ satisfies $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$. This identity, plus equivariance of δ implies

$$\delta(\mathbf{I}) = \delta(\mathbf{U}\mathbf{I}\mathbf{U}^T) = \mathbf{U}\delta(\mathbf{I})\mathbf{U}^T$$

for any orthogonal matrix $\mathbf{U}$. This implies that $\delta(\mathbf{I}) = c\mathbf{I}$ for some scalar $c > 0$. We therefore have that any equivariant estimator δ must satisfy

$$\delta(\mathbf{S}) = c\mathbf{S}$$

for some $c > 0$.

Since $\bar{\mathcal{G}}$ acts on the parameter space transitively, we should be able to obtain a UMRE estimator by identifying the value of c that minimizes the (constant) risk:

$$\begin{aligned} R(\Sigma, c\mathbf{S}) = R(\mathbf{I}, c\mathbf{S}) &= \mathrm{E}[\mathrm{tr}(c\mathbf{S}) - \log|c\mathbf{S}||\mathbf{I}] - p \\ &= c \times \mathrm{tr}(\mathrm{E}[\mathbf{S}|\mathbf{I}]) - p\log c - \mathrm{E}[\log|\mathbf{S}||\mathbf{I}] - p \\ &= c(n-1)p - p\log c + k. \end{aligned}$$

This is minimized in c by $c = 1/(n-1)$, and so the best equivariant estimator based on $\mathbf{S}$ (under this loss) is the unbiased estimator $\delta(\mathbf{S}) = \mathbf{S}/[n-1]$.

Example (binomial proportion):

$X \sim$ binary(n, θ), $\theta \in [0, 1]$.
$L(\theta, d) = (\theta - d)^2$
$\mathcal{G} = \{g_0(x), g_1(x)\}$, $g_0(x) = x$, $g_1(x) = n - x$.

Model: $X \sim$ binary$(n, \theta) \Rightarrow g_1(X) = n - X \sim$ binary$(n, 1-\theta)$, so the model is invariant.

Parameter space: $\bar{g}_1(\theta) = 1 - \theta$.

Decision space: If $d' = \tilde{g}_1(d) = 1 - d$

$$\begin{aligned} L(\bar{g}(\theta), \tilde{g}(d)) &= ([1-\theta] - [1-d])^2 \\ &= (\theta - d)^2 = L(\theta, d) \end{aligned}$$

The orbit of $\theta_0 \in \Theta$ is only $\{\theta_0, 1-\theta_0\}$. $\bar{\mathcal{G}}$ is not rich enough to be transitive - we might then expect that equivariant estimators are not generally constant risk. Let's check: An estimator is equivariant if

$$\begin{aligned} \delta(g_1 X) &= \tilde{g}_1 \delta(X) \\ \delta(n - X) &= 1 - \delta(X). \end{aligned}$$

The class of equivariant estimators is quite large, and includes (for example) the Bayes estimators under all priors symmetric about 1/2. These estimators do not have constant risk, and there is no best decision rule in this class. Invariance is not a big help in discriminating among procedures.

5 Procedures for finding UMRE estimators

Lehmann and Casella [1998] sections 3.1 and 3.3 present some techniques for finding UMREEs for location, scale and location-scale problems specifically, but they don't discuss how these techniques tie together or generalize. Berger [1985] section 6.5 discusses a general method for finding UMREEs, but doesn't go into much detail, or show many examples. We'll try to tie the material from both of these sources together.

5.1 Location models

Let $\boldsymbol{X} \sim P_\theta \in \mathcal{P} = \{p(\boldsymbol{x}|\theta) : p_0(x_1 - \theta, \ldots, x_n - \theta), \theta \in \Theta = \mathbb{R}\}$.

Here, p_0 is a known probability density on $\mathbb{R}^n$. Some possibilities include

- an i.i.d. model : $p_0(\boldsymbol{x}) = \prod_{i=1}^n f_0(x_i)$
- an independence model: $p_0(\boldsymbol{x}) = \prod_{i=1}^n f_i(x_i)$
- a dependence model: $p_0(\boldsymbol{x}) = \prod_{i=1}^n f_i(x_i|x_{i-1})$

Model invariance and induced groups:

$\boldsymbol{X} \sim P_\theta, \theta \in \Theta$.

$\boldsymbol{X}' = g\boldsymbol{X} = \boldsymbol{X} + a\mathbf{1},\ a \in \mathbb{R}$.

By the usual change of variables formula, the joint density of $\boldsymbol{X}$ is

$$\begin{aligned} p_{x'}(\boldsymbol{x}') &= p(\boldsymbol{x}(\boldsymbol{x}')|\theta) \times |d\boldsymbol{x}/d\boldsymbol{x}'| \\ &= p(x_1' - a, \ldots, x_n' - a|\theta) \\ &= p_0(x_1' - (\theta + a), \ldots, x_n' - (\theta + a)) \\ &= p(\boldsymbol{x}'|\theta + a) \in \mathcal{P}. \end{aligned}$$

It is also clear that such a transformation doesn't reduce the model, and so

- $\mathcal{P}$ is invariant under the additive group $\mathcal{G} = \{g : \boldsymbol{x} \to \boldsymbol{x} + a, a \in \mathbb{R}\}$;
- the induced group on Θ is $\bar{\mathcal{G}} = \{\bar{g} : \theta \to \theta + a, a \in \mathbb{R}\}$.

Note that $\bar{\mathcal{G}}$ is a transitive group on Θ.
Consider estimation of θ under a loss of the form

$$L(\theta, d) = \rho(d - \theta)$$

where ρ is an increasing function with $\rho(0) = 0$. It is straightforward to show that such an L is invariant, with

$$\tilde{\mathcal{G}} = \{\tilde{g} : d \to d + a, a \in \mathbb{R}\}.$$

Definition 7 (LC 1.2)**.** *Under the model $\mathcal{P}$ and loss L, the problem of estimating θ is said to be location invariant.*

For this problem, an equivariant estimator is one that satisfies

$$\begin{aligned} \delta(g\boldsymbol{x}) &= \tilde{g}\delta(\boldsymbol{x}) \\ \delta(\boldsymbol{x} + a\mathbf{1}) &= \delta(\boldsymbol{x}) + a. \end{aligned}$$

As $\bar{\mathcal{G}}$ is transitive, we already know from our general results that $R(\theta, \delta)$ is constant in θ. LC prove this result for this special case:

Theorem (LC 3.1.4)**.** *For the invariant location problem, the bias, variance and risk of any equivariant estimator are constant as a function of $\theta \in \mathbb{R}$.*

<u>Exercise:</u> Prove this theorem.

Thus the risk function of any estimator is determined by its risk at any value of θ. A convenient choice is $\theta = 0$. Finding the UMREE then amounts to

1. calculating $R(0, \delta)$ for each equivariant estimator δ,
2. selecting the δ that minimizes the risk.

5.2 Characterizing equivariance

Before we compare risk functions, we need to characterize a form for all equivariant estimators. The first thing to notice is that the difference between any two equivariant estimators is an *invariant* function of $\boldsymbol{x}$:

Let δ and δ_0 be equivariant. Then

$$\begin{aligned} u(\boldsymbol{x}) &\equiv \delta(\boldsymbol{x}) - \delta_0(\boldsymbol{x}) \\ u(g\boldsymbol{x}) &= \delta(g\boldsymbol{x}) - \delta_0(g\boldsymbol{x}) \\ &= \tilde{g}\delta(\boldsymbol{x}) - \tilde{g}\delta_0(\boldsymbol{x}) \quad \text{(by equivariance)} \\ &= \delta(\boldsymbol{x}) + a - \delta_0(\boldsymbol{x}) - a \\ &= \delta(\boldsymbol{x}) - \delta_0(\boldsymbol{x}) = u(\boldsymbol{x}). \end{aligned}$$

This leads to the following characterization of equivariant estimators:

Theorem (LC 3.1.6). *Let δ_0 be an equivariant estimator. Then δ is equivariant iff*

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) + u(\boldsymbol{x}),$$

for some $u(\boldsymbol{x})$ such that $u(g\boldsymbol{x}) = u(\boldsymbol{x})$ $\forall g \in \mathcal{G}$.

Proof. We have shown that if δ and δ_0 are equivariant, then their difference is invariant. Now suppose δ_0 is equivariant and u is invariant, and let $\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) + u(\boldsymbol{x})$. Then

$$\begin{aligned} \delta(\boldsymbol{x} + a\mathbf{1}) = \delta(gx) &= \delta_0(g\boldsymbol{x}) + u(g\boldsymbol{x}) \\ &= \tilde{g}\delta_0(\boldsymbol{x}) + u(\boldsymbol{x}) \\ &= \delta_0(\boldsymbol{x}) + a + u(\boldsymbol{x}) = \delta(\boldsymbol{x}) + a, \end{aligned}$$

and so such a δ is equivariant. □

Maximal invariants:

The next step to characterizing the location equivariant estimators is to characterize the invariant functions $u(\boldsymbol{x})$. Here is the lemma presented in LC:

Lemma (LC 3.1.7)**.** *A function $u(\boldsymbol{x})$ satisfies $u(\boldsymbol{x}+a) = u(\boldsymbol{x})$ iff*

- *for $n > 1$, it is a function of the differences $y_1 = x_1 - x_n, \ldots, y_{n-1} = x_{n-1} - x_n$;*
- *for $n = 1$, it is a constant.*

Proof. For $n = 1$, if it is constant it is invariant. If it is invariant, then $u(x) = u(x - x) = u(0)\ \forall x \in \mathbb{R}$, and so it is constant.

For $n > 1$

- Suppose u is a function of the differences, i.e. $u(\boldsymbol{x}) = h(\boldsymbol{y}(\boldsymbol{x}))$. Letting $\boldsymbol{x}' = \boldsymbol{x} + a$,

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

- Suppose u is invariant. Then

$$
\begin{aligned}
u(\boldsymbol{x}) &= u(\boldsymbol{x} - x_n) \\
&= u(x_1 - x_n, \ldots, x_{n-1} - x_n, 0) \\
&= h(\boldsymbol{y}(\boldsymbol{x})).
\end{aligned}
$$

□

We have shown that any function $u(\boldsymbol{x})$ invariant under the additive group must be a function of the differences. Of course, the differences themselves are invariant. We say that the differences $\boldsymbol{y}(\boldsymbol{x})$ are a *maximal invariant statistic*. More generally, we have the following definitions and result:

Let $\mathcal{G}$ be a group on $\mathcal{X}$.

Definition. *A function $u : \mathcal{X} \to \mathcal{U}$ is invariant if*

$$u(gx) = u(x) \ \forall g \in \mathcal{G}, x \in \mathcal{X}.$$

Definition. *A function $y : \mathcal{X} \to \mathcal{Y}$ is a maximal invariant if it is invariant and if for any two points $x_1, x_2 \in \mathcal{X}$,*

$$y(x_1) = y(x_2) \Rightarrow x_2 = gx_1 \text{ for some } g \in \mathcal{G}.$$

In other words,

- invariant functions are constant on orbits of $\mathcal{X}$.
- a maximal invariant function *identifies* the orbits of $\mathcal{X}$, i.e. takes on a different value on each orbit.

Exercise: Draw a picture of how a maximal invariant partitions the sample space. A simple example is the case of $\boldsymbol{x} \in \mathbb{R}^2$ and $\mathcal{G} = \{g : \boldsymbol{x} \to \boldsymbol{x} + (a, a), \ a \in \mathbb{R}\}$.

This last characteristic of maximal invariants implies the following:

Theorem 2. *A function $u(x)$ is invariant iff it is a function of a maximal invariant $y(x)$.*

Proof.
If u is a function of y then u is certainly invariant.
To go the other direction, we need to show that

if u is invariant, then

$y(x) = y(x')$ implies that $u(x) = u(x')$.

This would mean that $u(x)$ only changes as $y(x)$ changes.
So let u be invariant and suppose $y(x) = y(x')$.
Since y is maximal invariant, this means $x' = gx$ for some $g \in \mathcal{G}$.
But since u is invariant, we have $u(x') = u(gx) = u(x)$. □

Exercise (scale group):
Find a maximal invariant function of $\boldsymbol{x}$ for the multiplicative group $\mathcal{G} = \{g : \boldsymbol{x} \to a\boldsymbol{x}, a > 0\}$.

Returning to the location problem, we have shown the following:

- If δ_0 is any equivariant estimator, $\delta(\boldsymbol{x})$ is equivariant iff

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) + u(\boldsymbol{x})$$

  for some invariant $u(\boldsymbol{x})$;

- If $u(\boldsymbol{x})$ is invariant, then it is a function of the differences $\boldsymbol{y}$, which is a maximal invariant, so

$$u(\boldsymbol{x}) = v(\boldsymbol{y}(\boldsymbol{x})).$$

LC combine these results into the following theorem:

Theorem (LC 3.1.8). *If δ_0 is any equivariant estimator, then δ is equivariant iff*

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

where $v(\boldsymbol{y})$ is some function of the differences $y_1 = x_1 - x_n, \ldots, y_{n-1} = x_{n-1} - x_n$.

Example:
If $n = 1$ then the maximal invariant is constant, and so all equivariant estimators are of the form $\delta(x) = \delta_0(x) + c$, where c is constant in x and δ_0 is an arbitrary equivariant estimator. One such estimator is $\delta_0(x) = x$, giving our previous result that the equivariant estimators are $\delta(x) = x + c, c \in \mathbb{R}$.

We are now in a position to construct the minimum risk equivariant estimator. Recall, our objective is to minimize $R(\theta, \delta)$ among equivariant estimators. We'll do this under squared error loss (see LC for more general loss functions).

$$\begin{aligned} R(\theta, \delta) = R(0, \delta) &= \mathrm{E}_0[L(0, \delta_0(\boldsymbol{X}) - v(\boldsymbol{Y}))] \\ &= \mathrm{E}_0[(\delta_0(\boldsymbol{X}) - v(\boldsymbol{Y}))^2] \\ &= \mathrm{E}_0[\ \mathrm{E}_0[(\delta_0(\boldsymbol{X}) - v(\boldsymbol{Y}))^2 | \boldsymbol{Y}]\ ]. \end{aligned}$$

Now suppose for each $\boldsymbol{y}$, we choose $v^*(\boldsymbol{y})$ as the minimizer of the inner expectation:

$$v^*(\boldsymbol{y}) \text{ minimizes } \mathrm{E}_0[(\delta_0(\boldsymbol{X}) - v(\boldsymbol{y}))^2 | \boldsymbol{Y} = \boldsymbol{y}] \text{ over all functions } v.$$

Then clearly $\delta_0(\boldsymbol{X}) - v^*(\boldsymbol{Y})$ minimizes the unconditional risk, and is therefore the UMRE estimator. In a sense, what we have done here is find the best equivariant estimator on each orbit of $\mathcal{G}$ (recall, v is constant on orbits), and the maximal invariant $\boldsymbol{y}$ defines the orbits.

Can we find such a v^*?

$$\begin{aligned} \mathrm{E}_0[(\delta_0(\boldsymbol{X}) - v(\boldsymbol{y}))^2 | \boldsymbol{Y} = \boldsymbol{y}] &= \mathrm{E}_0[\delta_0^2(\boldsymbol{x}) - 2\delta_0(\boldsymbol{x})v(\boldsymbol{y}) + (\boldsymbol{y}) | \boldsymbol{Y} = \boldsymbol{y}] \\ &= \mathrm{E}_0[\delta_0^2(\boldsymbol{x})|\boldsymbol{y}] - 2v(\boldsymbol{y})\mathrm{E}_0[\delta_0(\boldsymbol{x})|\boldsymbol{y}] + v(\boldsymbol{y})^2. \end{aligned}$$

The unique minimize of the conditional expectation is therefore

$$v^*(\boldsymbol{y}) = \mathrm{E}_0[\delta_0(\boldsymbol{x})|\boldsymbol{y}],$$

and so the UMREE can be expressed as

$$\delta_0(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) - \mathrm{E}_0[\delta_0(\boldsymbol{x})|\boldsymbol{y}].$$

All that remains is to pick a convenient δ_0 for which the conditional expectation can be easily calculated.

LC give this derivation in a theorem that is applicable to more general loss functions, but the basic idea is the same.

Theorem (LC 3.1.10, LC 3.1.11)**.** *Let $L(\theta, d) = \rho(d - \theta)$, where ρ is convex and not monotone. If δ_0 is an equivariant estimator with finite risk, then the unique UMRE estimator is given by*

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) - v^*(\boldsymbol{y}),$$

where $v^*(\boldsymbol{y})$ is the minimizer of $\mathrm{E}_0[\rho(\delta_0(\boldsymbol{x}) - v(\boldsymbol{y}))|\boldsymbol{y}]$.

Corollary (LC 3.1.12)**.**

- *If $\rho(d - \theta) = (d - \theta)^2$, then $v^*(\boldsymbol{y}) = \mathrm{E}_0[\delta_0(\boldsymbol{X})|\boldsymbol{y}]$.*
- *If $\rho(d - \theta) = |d - \theta|$, then $v^*(\boldsymbol{y})$ is any median of $\delta_0(\boldsymbol{X})$ under $p_0(\boldsymbol{x}|\boldsymbol{y})$.*

5.3 The Pitman location estimator

The UMREE under squared error loss is

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x}) - \mathrm{E}_0[\delta_0(\boldsymbol{X})|\boldsymbol{y}],$$

where $\delta_0(\boldsymbol{x})$ is an arbitrary equivariant estimator. The choice of δ_0 will not affect δ, it will only affect our ability to calculate $\mathrm{E}_0[\delta_0(\boldsymbol{X})|\boldsymbol{y}]$. Let's try a very simple equivariant estimator:

$$\delta_0(\boldsymbol{x}) = x_n.$$

Note that

$$\begin{aligned} \delta_0(g\boldsymbol{x}) &= \delta_0(\boldsymbol{x} + a) \\ &= x_n + a \\ &= \delta_0(\boldsymbol{x}) + a = \tilde{g}\delta_0(\boldsymbol{x}), \end{aligned}$$

and so this estimator is equivariant. Now we just need to calculate the conditional expectation:

$$\mathrm{E}[X_n | X_1 - X_n, \ldots, X_{n-1} - X_n].$$

The joint density of $(X_1 - X_n, \dots, X_{n-1} - X_n, X_n) = (Y_1, \dots, Y_{n-1}, X_n)$ can be found from the change of variables formula:

$$p_{y,x_n}(y_1, \dots, y_{n-1}, x_n) = p_{\boldsymbol{x}}(x_1(\boldsymbol{y}, x_n), \dots, x_{n-1}(\boldsymbol{y}, x_n), x_n)|d(\boldsymbol{y}, x_n)/d\boldsymbol{x}|.$$

Now

- $|d(\boldsymbol{y}, x_n)/d\boldsymbol{x}| = 1$
- $x_i = y_i + x_n, \; i = 1, \dots n-1.$

so we have

$$p_{y,x_n}(y_1, \dots, y_{n-1}, x_n) = p_0(y_1 + x_n, \dots, y_{n-1} + x_n, x_n).$$

The conditional density for X_n given $Y_1, \dots, Y_{n-1}$ is

$$p_{x_n|\boldsymbol{y}}(x_n|\boldsymbol{y}) = \frac{p_0(y_1 + x_n, \dots, y_{n-1} + x_n, x_n)}{\int p_0(y_1 + x, \dots, y_{n-1} + x, x) \; dx},$$

and so the desired conditional expectation is

$$\mathrm{E}[X_n|Y_1, \dots, Y_{n-1}] = \frac{\int x p_0(y_1 + x, \dots, y_{n-1} + x, x) \; dx}{\int p_0(y_1 + x, \dots, y_{n-1} + x, x) \; dx}.$$

Now do a change of variables: Let $x = x_n - \theta$, where x_n is the observed value of X_n, and θ is the variable of integration:

$$\mathrm{E}[X_n|Y_1, \dots, Y_{n-1}] = \frac{\int (x_n - \theta) p_0(y_1 + x_n - \theta, \dots, y_{n-1} + x_n - \theta, x_n - \theta) \; d\theta}{\int p_0(y_1 + x_n - \theta, \dots, y_{n-1} + x_n - \theta, x_n - \theta) \; d\theta}.$$

Now recall $y_i = x_i - x_n$ for $i = 1, \dots, n-1$:

$$\begin{aligned}
\mathrm{E}[X_n|\boldsymbol{y}] &= \frac{\int (x_n - \theta) p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta}{\int p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta} \\
&= x_n - \frac{\int \theta p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta}{\int p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta} \\
\delta(\boldsymbol{x}) &= x_n - \mathrm{E}[X_n|\boldsymbol{y}] \\
&= \frac{\int \theta p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta}{\int p_0(x_1 - \theta, \dots, x_n - \theta) \; d\theta}.
\end{aligned}$$

This is known as the Pitman estimator of θ. Note that it is equal to the generalized Bayes estimator under the prior measure $\pi(\theta) \propto 1, \theta \in \mathbb{R}$. Under this "prior,"

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &= \frac{\pi(\theta)p_0(x_1-\theta,\ldots,x_n-\theta)}{\int \pi(\theta)p_0(x_1-\theta,\ldots,x_n-\theta)\ d\theta} \\
&= \frac{p_0(x_1-\theta,\ldots,x_n-\theta)}{\int p_0(x_1-\theta,\ldots,x_n-\theta)\ d\theta}, \text{ so} \\
\mathrm{E}_\pi[\theta|\boldsymbol{x}] &= \frac{\int \theta p_0(x_1-\theta,\ldots,x_n-\theta)\ d\theta}{\int p_0(x_1-\theta,\ldots,x_n-\theta)\ d\theta}.
\end{aligned}$$

Example (normal model):

$X_1,\ldots,X_n \sim$ i.i.d. $N(\theta,\sigma^2)$, σ^2 known.

Let's find the "posterior" that corresponds to the equivariant estimator:

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto p_0(\boldsymbol{x}-\theta\mathbf{1}) \\
&= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2}\exp(-(x_i-\theta)^2/[2\sigma^2]) \\
&\propto_\theta \exp\{\theta\bar{x}n/\sigma^2 - \tfrac{1}{2}\theta^2 n/\sigma^2\} \\
&\propto_\theta \mathrm{dnorm}(\theta,\bar{x},\sqrt{\sigma^2/n}).
\end{aligned}$$

So the UMREE under squared error loss is the generalized Bayes estimator based on this posterior, $\delta(\boldsymbol{x}) = \bar{x}$, the same as the UMVUE. Note that since the UMREE doesn't depend on σ^2, it must also be UMREE in the case that σ^2 is unknown.

Example (exponential model):

Let $X_1,\ldots,X_n \sim$ i.i.d. exponential(θ,b), with b known. In this case,

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &\propto p_0(\boldsymbol{x}-\theta\mathbf{1}) \\
&= \prod_{i=1}^n 1(x_i>\theta)\exp\{-(x_i-\theta)/b\}/b \\
&\propto_\theta 1(x_{(1)}>\theta)e^{n\theta/b} \\
\mathrm{E}[\theta|\boldsymbol{x}] &= \frac{\int_{-\infty}^{x_{(1)}} \theta e^{n\theta/b}\ d\theta}{\int_{-\infty}^{x_{(1)}} e^{n\theta/b}\ d\theta}
\end{aligned}$$

Using integration by parts, the numerator is

$$\tfrac{b}{n}\theta e^{n\theta/b} - \tfrac{b^2}{n^2}e^{n\theta/b}\big|_{-\infty}^{x_{(1)}} = \tfrac{b}{n}x_{(1)}e^{nx_{(1)}/b} - \tfrac{b^2}{n^2}e^{nx_{(1)}/b},$$

whereas the denominator is

$$\tfrac{b}{n}e^{n\theta/b}\big|_{-\infty}^{x_{(1)}} = \tfrac{b}{n}e^{nx_{(1)}/b},$$

giving the UMREE as $\delta(\boldsymbol{x}) = x_{(1)} - b/n$, the same as the UMVUE. However, note that the UMREE depends on the scale parameter b. Thus there won't exist an UMREE for this group when b is unknown.

Example (uniform):
Let $X_1, \ldots, X_n \sim$ i.i.d. uniform$(\theta - 1/2, \theta + 1/2)$.
The UMREE is $(X_{(1)} + X_{(n)})/2$.
There is no UMVUE (see LC Example 2.1.9).

5.4 Scale models

Invariant scale problem
Let $X_1, \ldots, X_n \sim$ i.i.d. $P_\sigma \in \mathcal{P} = \{p_1(x_1/\sigma, \ldots, x_n/\sigma)/\sigma^n, \sigma > 0\}$.
Using the usual change of variables formula, it can be shown that this model is invariant under transformations

$$\mathcal{G} = \{g : \boldsymbol{x} \to b\boldsymbol{x}, b > 0\},$$

which induce the following group of transformations on the parameter space:

$$\bar{\mathcal{G}} = \{\bar{g} : \sigma \to b\sigma, b > 0\}.$$

Consider estimation of σ^r under a loss of the form

$$L(\sigma, d) = \gamma(d/\sigma^r).$$

Examples:

- standardized power loss: $L(\sigma, d) = |d - \sigma^r|^p/\sigma^{pr} = |\frac{d}{\sigma^r} - 1|^p$;
- Stein's loss: $L(\sigma, d) = (d/\sigma^r) - \log(d/\sigma^r) - 1$.

Note that

- $\lim_{d\to\infty} L(\sigma, d) = \infty, \lim_{d\to 0} L(\sigma, d) = 1$ for power loss, but
- $\lim_{d\to\infty} L(\sigma, d) = \infty, \lim_{d\to 0} L(\sigma, d) = \infty$ for Stein's loss.

Any of these loss functions are invariant, with the induced group on D being

$$\tilde{\mathcal{G}} = \{\tilde{g} : d \to b^r d, d > 0\}.$$

Equivariant estimation

An equivariant estimator for this problem is one that satisfies

$$\delta(g\boldsymbol{x}) = \tilde{g}\delta(\boldsymbol{x})$$
$$\delta(b\boldsymbol{x}) = b^r\delta(\boldsymbol{x}).$$

Notice that $\bar{\mathcal{G}}$ and $\tilde{\mathcal{G}}$ act transitively and commutatively on Θ and D.
Exercise: Prove the following theorem:

Theorem. *Let δ_0 be an equivariant estimator. Then δ is equivariant iff*

$$\delta(\boldsymbol{x}) = u(\boldsymbol{x})\delta_0(\boldsymbol{x}),$$

for some $u(\boldsymbol{x})$ such that $u(g\boldsymbol{x}) = u(\boldsymbol{x})\ \forall g \in \mathcal{G}$.

Now let's characterize the invariant functions $u(\boldsymbol{x})$. Recall our theorem:

Theorem. *Any invariant function is a function of the maximal invariant.*

What is the maximal invariant for this multiplicative group?
What function of $\boldsymbol{x}$ contains all of the information, except the "scale"?

Theorem. *For $\mathcal{G} = \{g : \boldsymbol{x} \to b\boldsymbol{x}, b > 0\}$ acting on $\boldsymbol{x} \in \mathbb{R}^n$, a maximal invariant is*

$$\boldsymbol{z}(\boldsymbol{x}) = (x_1/x_n, \dots, x_{n-1}/x_n, x_n/|x_n|).$$

Proof. Clearly $\boldsymbol{z}(b\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x})$. Now let's show it is maximal, i.e.

$$\boldsymbol{z}(\boldsymbol{x}') = \boldsymbol{z}(\boldsymbol{x}) \Rightarrow \boldsymbol{x}' = b\boldsymbol{x} \text{ for some } b > 0.$$

So suppose $\boldsymbol{z}(\boldsymbol{x}') = \boldsymbol{z}(\boldsymbol{x})$.
Then $z_i' = x_i'/x_n' = x_i/x_n = z_i,\ i = 1, \dots, n-1$.
This implies $x_i' = (x_n'/x_n)x_i = bx_i,\ i = 1, \dots, n-1$.
Note that $b = (x_n'/x_n) > 0$ because $z_n' = x_n'/|x_n'| = x_n/|x_n| = z_n$.
Finally, we trivially have $x_n' = (x_n'/x_n)x_n = bx_n$.
Thus $\boldsymbol{x}' = b\boldsymbol{x}$. □

Our characterization of equivariant estimators is therefore as follows:

Theorem (LC Theorem 3.3.1)**.** *Let δ_0 be any equivariant estimator of σ^r. Then δ is equivariant iff there exists a function $w(\boldsymbol{z})$ such that*

$$\delta(\boldsymbol{x}) = \delta_0(\boldsymbol{x})/w(\boldsymbol{z}).$$

Based on this representation, we can identify the UMREE via a conditioning argument as before. For simplicity, let's consider estimation of σ^r with scaled squared-error loss ($p = 2$).

$$\begin{aligned} R(\sigma^2, \delta) = R(1, \delta) &= \mathrm{E}_1 L(1, \delta) \\ &= \mathrm{E}_1[(\delta_0(\boldsymbol{x})/w(\boldsymbol{z}) - 1)^2] \\ &= \mathrm{E}_1[\mathrm{E}_1[(\delta_0(\boldsymbol{x})/w(\boldsymbol{z}) - 1)^2|\boldsymbol{z}]] \end{aligned}$$

The optimal $w(\boldsymbol{z})$ can be found by having $w(\boldsymbol{z})$ minimize the conditional expectation for each $\boldsymbol{z}$.

$$\begin{aligned} \mathrm{E}_1[(\delta_0(\boldsymbol{x})/w(\boldsymbol{z}) - 1)^2|\boldsymbol{z}] &= \mathrm{E}_1[\tfrac{\delta^2}{w^2} - 2\tfrac{\delta_0}{w} + 1|\boldsymbol{z}] \\ &= \mathrm{E}_1[\delta_0^2|\boldsymbol{z}]/w^2 - 2\mathrm{E}_1[\delta_0|\boldsymbol{z}]/w + 1 \end{aligned}$$

Taking derivatives, the minimum in w satisfies

$$-2\mathrm{E}_1[\delta_0^2|\boldsymbol{z}]/w^3 + 2\mathrm{E}_1[\delta_0|\boldsymbol{z}]/w^2 = 0$$
$$w(\boldsymbol{z}) = \mathrm{E}_1[\delta_0^2|\boldsymbol{z}]/\mathrm{E}_1[\delta_0|\boldsymbol{z}].$$

Therefore, the UMREE under scaled squared error satisfies

$$\delta(\boldsymbol{x}) = \frac{\delta_0(\boldsymbol{x})\mathrm{E}_1[\delta_0(\boldsymbol{x})|\boldsymbol{z}]}{\mathrm{E}_1[\delta_0^2(\boldsymbol{x})|\boldsymbol{z}]}.$$

Similar calculations show that the UMREE under Stein's loss is

$$\delta(\boldsymbol{x}) = \frac{\delta_0(\boldsymbol{x})}{\mathrm{E}_1[\delta_0(\boldsymbol{x})|\boldsymbol{z}]}.$$

In either case, the remaining task is to choose $\delta_0(\boldsymbol{x})$ and calculate its conditional moments, given $\boldsymbol{z}$.

Example (normal variance, known mean):

$X_1, \ldots, X_n \sim$ i.i.d. $N(0, \sigma^2)$,

A good strategy for picking $\delta_0(\boldsymbol{x})$ is to choose

- a really simple estimator (so conditional calculations are easy), or
- a really good estimator (so maybe the adjustment to δ_0 will be simple).

Consider $\delta_0(\boldsymbol{x}) = \sum x_i^2$, which is equivariant.

$\sum x_i^2$ is a complete sufficient statistic,

$\boldsymbol{z}(\boldsymbol{x})$ is ancillary (has a distribution that doesn't depend on σ^2)

$\sum x_i^2$ is independent of $\boldsymbol{z}(\boldsymbol{x})$ by Basu's theorem.

Thus $\mathrm{E}_1[\delta|\boldsymbol{z}] = \mathrm{E}_1[\delta]$, $\mathrm{E}_1[\delta^2|\boldsymbol{z}] = \mathrm{E}_1[\delta^2]$,

Stein's loss:

The estimator under Stein's loss is

$$\frac{\delta_0(\boldsymbol{x})}{\mathrm{E}_1[\delta_0(\boldsymbol{x})|\boldsymbol{z}]} = \frac{\delta_0(\boldsymbol{x})}{\mathrm{E}_1[\delta_0(\boldsymbol{x})]} = \frac{\sum x_i^2}{n},$$

which is also the UMVUE.

Scaled squared error loss:
The estimator under scaled squared error loss is

$$\frac{\delta_0(\boldsymbol{x})\mathrm{E}_1[\delta_0(\boldsymbol{x})]}{\mathrm{E}_1[\delta_0^2(\boldsymbol{x})]}.$$

The expectations can be calculated from the moments of the gamma distribution, giving the UMREE as

$$\delta(\boldsymbol{x}) = \frac{\sum x_i^2}{n+2}$$

Bayesian representation:
The conditional expectations are more tedious in the absence of a complete sufficient statistic. For such cases, let's pick a simple equivariant estimator to begin with:

$$\delta_0(\boldsymbol{x}) = |x_n|^r.$$

The UMREE is based on conditional moments of $|X_n|$ given $\boldsymbol{z}$. To calculate these moments, we need the conditional density of $|X_n|$ given $\boldsymbol{z}$. To make notation easier, let's denote the maximal invariant as

$$(\boldsymbol{z}, s) = (x_1/x_n, \ldots, x_{n-1}/x_n, x_n/|x_n|).$$

Note that s is the sign of x_n.

$$\begin{aligned}
\Pr(|X_n| \leq t|\boldsymbol{z}, s) &= \frac{\int_{-t}^{t} p_{\boldsymbol{z},x_n,s}(\boldsymbol{z}, x, s)\ dx}{\int_{-\infty}^{\infty} p_{\boldsymbol{z},x_n,s}(\boldsymbol{z}, x, s)\ dx} \\
&= \frac{\int_{-t}^{t} p_{\boldsymbol{z},x_n}(\boldsymbol{z}, x) p_{s|x_n}(s|x)\ dx}{\int_{-\infty}^{\infty} p_{\boldsymbol{z},x_n}(\boldsymbol{z}, x) p_{s|x_n}(s|x)\ dx}.
\end{aligned}$$

Now note that

$$\begin{aligned}
p_{s|x_n}(s|x) &= 1(x > 0) \text{ if } s = 1 \\
&= 1(x < 0) \text{ if } s = -1\ .
\end{aligned}$$

Let's suppose the observed value of x_n is positive, i.e. $s = 1$. In this case,

$$\Pr(|X_n| \leq t|\boldsymbol{z}, s) = \frac{\int_0^t p_{\boldsymbol{z},x_n}(\boldsymbol{z}, x)\ dx}{\int_0^\infty p_{\boldsymbol{z},x_n}(\boldsymbol{z}, x)\ dx}$$

Therefore, when $s = 1$ the conditional density of $T = |X_n|$ is

$$p_{t|\boldsymbol{z},s}(t|\boldsymbol{z}, s = 1) = \frac{p_{\boldsymbol{z},x_n}(\boldsymbol{z}, t)}{\int_0^\infty p_{\boldsymbol{z},x_n}(\boldsymbol{z}, t)\ dt},$$

where we have replaced x by t in the denominator. Now we need to find $p_{\boldsymbol{z},x_n}(\boldsymbol{z}, x_n)$:

$$\begin{aligned} p_{\boldsymbol{z},x_n}(z_1, \ldots, z_{n-1}, x_n) &= p_1(x_1(\boldsymbol{z}, x_n), \ldots, x_n(\boldsymbol{z}, x_n) \times |d\boldsymbol{x}/d(\boldsymbol{z}, x_n)| \\ &= p_1(z_1 x_n, \ldots, z_{n-1}x_n, x_n)|x_n|^{n-1}. \end{aligned}$$

The conditional density of t is then

$$p_{t|\boldsymbol{z},s}(t|\boldsymbol{z}, s = 1) = \frac{p_1(z_1 t, \ldots, z_n t, t)t^{n-1}}{\int_0^\infty p(z_1 t, \ldots, z_n t, t)t^{n-1}\ dt},$$

The conditional expectation of $|X_n|^r$ given $\boldsymbol{z}$ and $s = 1$ is then

$$\mathrm{E}_1[|X_n|^r|\boldsymbol{z}, s = 1] = \frac{\int_0^\infty t^r p(z_1 t, \ldots, z_n t, t)t^{n-1}\ dt}{\int_0^\infty p(z_1 t, \ldots, z_n t, t)t^{n-1}\ dt}$$

That's it! Except this result is hard to interpret. Let's make the result easier to understand with a change variables: Letting $t = x_n v$, where x_n is the observed value of X_n, we have

$$\mathrm{E}_1[|X_n|^r|\boldsymbol{z}, s = 1] = \frac{\int_0^\infty x_n^r v^r p(z_1 x_n v, \ldots, z_n x_n v, x_n v)x_n^{n-1}v^{n-1}|x_n|\ dv}{\int_0^\infty p(z_1 x_n v, \ldots, z_n x_n v, x_n v)x_n^{n-1}v^{n-1}|x_n|\ dv}$$

Recalling $z_i x_n = x_i,\ i = 1, \ldots, n-1$, and canceling some x_n's gives

$$\mathrm{E}_1[|X_n|^r|\boldsymbol{z}, s = 1] = \frac{|x_n|^r \int_0^\infty v^r p(x_1 v, \ldots, x_n v)v^{n-1}\ dv}{\int_0^\infty p(x_1 v, \ldots, x_n v)v^{n-1}\ dv}.$$

Similar calculations show that the result is the same for $s = -1$. From this we can obtain the result in LC 3.19: Under scaled squared error loss, the UMREE is given by

$$\begin{aligned}
\delta(\boldsymbol{x}) &= |x_n|^r \frac{\mathrm{E}_1[|X_n|^r|\boldsymbol{z}, s]}{\mathrm{E}_1[|X_n|^{2r}|\boldsymbol{z}, s]} \\
&= |x_n|^r \frac{|x_n|^r \int_0^\infty v^r p(x_1 v, \ldots, x_n v) v^{n-1}\, dv}{|x_n|^{2r} \int_0^\infty v^{2r} p(x_1 v, \ldots, x_n v) v^{n-1}\, dv} \\
&= \frac{\int_0^\infty v^{n+r-1} p(x_1 v, \ldots, x_n v)\, dv}{\int_0^\infty v^{n+2r-1} p(x_1 v, \ldots, x_n v)\, dv}.
\end{aligned}$$

This is known as Pitman's estimator of σ^r. As you may be able to tell, this estimator is related to a generalized Bayes estimator of σ. To see this consider the change of variables $\sigma = 1/v$, $|dv/d\sigma| = \sigma^{-2}$:

$$\begin{aligned}
\delta(\boldsymbol{x}) &= \frac{\int_0^\infty v^{n+r-1} p(x_1 v, \ldots, x_n v)\, dv}{\int_0^\infty v^{n+2r-1} p(x_1 v, \ldots, x_n v)\, dv} \\
&= \frac{\int_0^\infty \sigma^{-n-r+1} p(x_1/\sigma, \ldots, x_n/\sigma) \sigma^{-2}\, d\sigma}{\int_0^\infty \sigma^{-n-2r+1} p(x_1/\sigma, \ldots, x_n/\sigma) \sigma^{-2}\, d\sigma} \\
&= \frac{\int_0^\infty \sigma^{-r} [p(x_1/\sigma, \ldots, x_n/\sigma)/\sigma^n] \sigma^{-1}\, d\sigma}{\int_0^\infty \sigma^{-2r} [p(x_1/\sigma, \ldots, x_n/\sigma)/\sigma^n] \sigma^{-1}\, d\sigma} \\
&= \frac{\int_0^\infty \sigma^{-r} p(\boldsymbol{x}|\sigma)\pi(\sigma) d\sigma}{\int_0^\infty p(\boldsymbol{x}|\sigma)\pi(\sigma) d\sigma} \times \left(\frac{\int_0^\infty \sigma^{-2r} p(\boldsymbol{x}|\sigma)\pi(\sigma) d\sigma}{\int_0^\infty p(\boldsymbol{x}|\sigma)\pi(\sigma) d\sigma} \right)^{-1} \\
&= \mathrm{E}[\sigma^{-r}|\boldsymbol{x}]/\mathrm{E}[\sigma^{-2r}|\boldsymbol{x}],
\end{aligned}$$

where $\pi(\sigma)$ is the (improper) prior $\pi(\sigma) = 1/\sigma$. To see that this is in fact the Bayes estimator under this prior and scaled squared error loss, recall the Bayes estimator $\delta_\pi(\boldsymbol{x})$ is obtained as the minimizer of the posterior risk:

$$\begin{aligned}
R(\pi, d|\boldsymbol{x} &= \mathrm{E}[(d - \sigma^r)^2/\sigma^{2r}|\boldsymbol{x}] \\
&= \mathrm{E}[d^2/\sigma^{2r} - d/\sigma^r + 1|\boldsymbol{x}] \\
&= d^2 \mathrm{E}[\sigma^{-2r}|\boldsymbol{x}] - d\mathrm{E}[\sigma^{-r}|\boldsymbol{x}] + 1.
\end{aligned}$$

Taking derivatives, we see that this is minimized by

$$d = \frac{\mathrm{E}[\sigma^{-r}|\boldsymbol{x}]}{\mathrm{E}[\sigma^{-2r}|\boldsymbol{x}]}.$$

Thus the UMREE is equal to the Bayes estimator under the improper prior $\pi(\sigma) = 1/\sigma$.

6 Location-scale models via invariant measure

In both of the last two examples, the UMREE could be derived as a generalized Bayes estimator under an improper prior. This result holds more generally: If the induced group $\bar{\mathcal{G}}$ on Θ is transitive, then there is a "formula" for the UMREE, and it can be obtained by turning the Bayesian crank of posterior risk minimization.

Theorem (Eaton [1989])**.** *If $\mathcal{G}$ acts properly on $\mathcal{X}$ and $\bar{\mathcal{G}}$ acts transitively on Θ, the UMREE is given by*

$$\delta(x) = \arg\min_d \int_{\bar{\mathcal{G}}} L(d, \bar{g}\theta_0) p(x|\bar{g}\theta_0) \mu_r(d\bar{g})$$

where μ_r is the right-invariant measure on $\bar{\mathcal{G}}$ and θ_0 is any point in Θ.

Before discussing what right-invariant measure is, we first consider the similarities between the UMREE δ given above and posterior risk minimization: Having seen x, the best equivariant decision is obtained as the minimizer of some integral. Also note that in the above integral,

- as $\bar{g}$ ranges over $\bar{\mathcal{G}}$, $\bar{g}\theta_0$ ranges over Θ;
- as $\bar{\mathcal{G}}$ is transitive, the range of $\bar{g}\theta_0$ is all of Θ.

Therefore, the integral over $\bar{\mathcal{G}}$ can essentially be viewed as an integral over Θ. To be more precise, define the right-invariant prior as follows:

Definition (Right-invariant prior measure)**.** *The right invariant prior measure corresponding to a group $\bar{\mathcal{G}}$ on Θ is defined as*

$$\pi_r(A) = \mu_r(\bar{g} : \bar{g}\theta_0 \in A) = \int 1(\bar{g}\theta_0 \in A)\ \mu_r(d\bar{g})\ \forall A \in \sigma(\Theta),$$

where θ_0 is any point in Θ and μ_r is the right-invariant measure on $\bar{\mathcal{G}}$.

Conceptually, π_r is the distribution of $\bar{g}\theta_0$ when $\bar{g}$ is selected randomly from μ_r. As we will soon see, the choice of θ_0 has no effect on π_r (this is due to the transitivity of the group and the right invariance of μ_r). This invariant prior allows us to view the UMREE as a Bayesian procedure.

Corollary. *Let $\bar{\mathcal{G}}$ act transitively on Θ. Then the UMREE is given by*

$$\delta(x) = \arg\min_d R(\pi_r, d|x) = \int_\Theta L(d,\theta)\pi_r(d\theta|x)$$

where

$$\pi_r(\theta|x) = \frac{p(x|\theta)\pi_r(\theta)}{\int p(x|\theta')\ \pi_r(d\theta')}$$

is the posterior density of θ based on the right-invariant prior π_r on Θ.

Proof. The UMREE is given by the minimizer of

$$\int_{\bar{\mathcal{G}}} L(d,\bar{g}\theta_0)p(x|\bar{g}\theta_0)\mu_r(d\bar{g}).$$

Letting $\theta = \bar{g}\theta_0$, this integral is the same as

$$\int_\Theta L(d,\theta)p(x|\theta)\pi_r(d\theta)$$

by the definition of π_r, and is proportional as a function of θ to the posterior risk. □

6.1 Invariant measure

To make use of these results we need to be able to obtain the right-invariant measure for an invariant estimation problem. The right-invariant measure μ_r on $\bar{\mathcal{G}}$ is a measure such that for any $g_0 \in \bar{\mathcal{G}}$ and measurable $A \subset \bar{\mathcal{G}}$

$$\mu_r(A\bar{g}_0^{-1}) = \mu_r(A).$$

The easiest way (for me) to think about this is in the case that μ_r is a probability measure, and we are sampling a $\bar{g}$ from $\bar{\mathcal{G}}$ according to μ_r. In this situation, for a right-invariant (probability) measure μ_r

$$\Pr(\bar{g} \in A) = \mu_r(A) = \mu_r(A\bar{g}_0^{-1}) = \Pr(\bar{g}\bar{g}_0 \in A),$$

that is, the probability that $\bar{g}$ is in A is the same as the probability that $\bar{g}\bar{g}_0$ is in A.

Example (additive group):
Let $\bar{\mathcal{G}} = \{\bar{g} : \theta \to \theta + a, a \in \mathbb{R}\}$. Then

- if $\bar{g} : \theta \to \theta + a$ and $\bar{g}_0 : \theta \to \theta + a_0$ then $\bar{g}\bar{g}_0 : \theta \to \theta + (a + a0)$;
- $\bar{\mathcal{G}}$ is isomorphic to $\mathbb{R}$.

μ_r is right-invariant on $\mathbb{R}$ means

$$\mu_r(A - a_0) = \mu_r(A) \ \ \forall a_0 \in \mathbb{R}, A \subset \mathcal{B}(\mathbb{R})$$

Clearly Lebesgue measure on $\mathbb{R}$ is right-invariant.

The induced prior π_r in this case is given by the distribution of $\theta_0 + a$ where $a \sim \mu_L$. Clearly, $\pi_r = \mu_L = \mu_r$.

Example (multiplicative group):
Let $\bar{\mathcal{G}} = \{\bar{g} : \theta \to a\theta, a > 0\}$. Then

- if $\bar{g} : \theta \to a\theta$ and $\bar{g}_0 : \theta \to a_0\theta$ then $\bar{g}\bar{g}_0 : \theta \to aa_0\theta$;
- $\bar{\mathcal{G}}$ is isomorphic to $\mathbb{R}^+$.

μ_r is right-invariant on $\mathbb{R}$ means

$$\mu_r(A/a_0) = \mu_r(A) \ \ \forall a_0 \in \mathbb{R}, A \subset \mathcal{B}(\mathbb{R})$$

Lebesgue measure on $\mathbb{R}^+$ is not right-invariant. For example, letting $A = (0, c)$ we have $\mu_L(A/a_0) = \mu_L(A)/a_0 \neq \mu_L(A)$ unless $a_0 = 1$. However, consider the measure μ with density $1/a$ with respect to Lebesgue measure μ_L. Pick any $B = (b_1, b_2) \subset \mathbb{R}^+$, and let $a' = aa_0$ (so $a = a'/a_0$). Then

$$\begin{aligned}
\mu(B/a_0) &= \int_{b_1/a_0}^{b_2/a_0} \tfrac{1}{a} \ \mu_L(da) \\
&= \int_{b_1}^{b_2} \tfrac{1}{a'/a_0} |\tfrac{da}{da'}| \mu_L(da) \\
&= \int_{b_1}^{b_2} \tfrac{a_0}{a'} \tfrac{1}{a_0} \ \mu_L(da') \\
&= \int_{b_1}^{b_2} \tfrac{1}{a'} \mu_L(da') = \mu(B),
\end{aligned}$$

so μ is right-invariant (you can use a $\pi - \lambda$ argument to show that $\mu(B/a_0) = \mu(B)$ for any Borel set). The relabeling μ as μ_r, we have for any $B \in \mathcal{B}(\mathbb{R})$

$$\mu_r(B) = \int_B a^{-1} \ \mu_L(da).$$

The induced prior π_r in this case is given by the distribution of $a\theta_0$ where $a \sim \mu_r$. by the invariance of μ_r, we have $\pi_r = \mu_r$.

We defined the induced prior π_r by $\pi_r(A) = \mu_r(\bar{g} : \bar{g}\theta_0 \in A)$, and claimed that π_r didn't depend on the value of θ_0. Let's make sure this is correct. Define two measures based on μ_r as follows:

$$\begin{aligned}
\pi_0(A) &= \mu_r(\{\bar{g} : \bar{g}\theta_0 \in A\}) \\
\pi_1(A) &= \mu_r(\{\bar{g} : \bar{g}\theta_1 \in A\}).
\end{aligned}$$

The claim is that these two measures are equal regardless of what the values of θ_0, θ_1 are. To see this, note that $\bar{\mathcal{G}}$ is transitive, and so there exists an $\bar{h} \in \bar{\mathcal{G}}$ such that $\theta_1 = \bar{h}\theta_0$. Therefore,

$$\begin{aligned}\{\bar{g} : \bar{g}\theta_1 \in A\} &= \{\bar{g} : \bar{g}\bar{h}\theta_0 \in A\} \\ &= \{\bar{g} : \bar{g}\theta_0 \in A\}\bar{h}^{-1}\end{aligned}$$

since if $\bar{g}\theta_0 \in A$, then $\bar{g}\bar{h}^{-1}$ is a transformation such that $\bar{g}\bar{h}^{-1}\bar{h}\theta_0 = \bar{g}\theta_0 \in A$. Now by the right invariance of μ_r, we have

$$\begin{aligned}\pi_1(A) &= \mu_r(\{\bar{g} : \bar{g}\theta_0 \in A\}\bar{h}^{-1}) \\ &= \mu_r(\{\bar{g} : \bar{g}\theta_0 \in A\}) = \pi_0(A).\end{aligned}$$

6.2 Equivariant estimation for location-scale families

Now we consider location scale models generated by a known density p_{01} on $\mathbb{R}^p$.

$$\mathcal{P} = \{p(\boldsymbol{x}|\mu,\sigma) = p_{01}(\boldsymbol{x} - \mu\mathbf{1})/\sigma)/\sigma^p : \{\mu,\sigma\} \in \mathbb{R} \times \mathbb{R}^+\}.$$

You can check that such models are invariant under functions of the form

$$\begin{aligned}g &: \boldsymbol{x} \to a\boldsymbol{x} + b \\ \bar{g} &: (\mu,\sigma) \to (a\mu + b, a\sigma)\end{aligned}$$

for $a > 0$, $b \in \mathbb{R}$.

A technicality: The group $\mathcal{G} = \{g : \mathbf{x} \to a\boldsymbol{x} + b,\ a > 0, b \in \mathbb{R}\}$ does not act properly on $\mathcal{X} = \mathbb{R}$. However, it does act properly on the subset of $\mathbb{R}$ for which $\sum(x_i - \bar{x})^2 > 0$. If this happens with probability 1 under p_{01}, then the above theorem will still work.

To apply the theorem, we need to find the right-invariant prior μ_r on $\bar{\mathcal{G}}$ and the induced right-invariant prior π_r on $\Theta = \{(\mu,\sigma) : \mu \in \mathbb{R}, \sigma > 0\}$. This is not too hard:

Lemma. *The right-invariant measure on* $\bar{G} = \{\bar{g} : (\mu, \sigma) \to (a\mu + b, a\sigma)\}$ *satisfies*

$$\mu_r(A \times B) = \int_A \int_B a^{-1} \, db \, da,$$

so that μ_r *has density* $1/a$ *with respect to Lebesgue measure on* $\mathbb{R} \times \mathbb{R}^+$.

Now recall the induced right-invariant prior π_r on (μ, σ) is the measure of $\bar{g}\theta_0$ when $\bar{g} \sim \mu_r$, for any θ_0. Taking $\theta_0 = (0, 1)$ and indexing $g : \boldsymbol{x} \to a\boldsymbol{x} + b\mathbf{1}$, we have that

$$(\mu, \sigma) \sim \pi_r \Leftrightarrow (\mu, \sigma) \stackrel{d}{=} \bar{g}_{ab}((0, 1)) = (b, a).$$

In other words, the invariant prior for (μ, σ) is the same as the invariant measure of (b, a).

Corollary. *The right-invariant prior on* $\Theta = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$ *has density* $\pi(\mu, \sigma) = 1/\sigma$ *with respect to Lebesgue measure on* $\mathbb{R} \times \mathbb{R}^+$.

Alternatively, the prior density in terms of (μ, σ^2) is given by

$$\pi(\mu, \sigma^2) = 1/\sigma^2.$$

We are now in a position to obtain the best equivariant estimator of any function of (μ, σ^2) under an invariant loss. Via the theorem, the best equivariant estimator is the minimizer of he posterior risk under the right-invariant prior.
Example (normal model): Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2), \ \mu \in \mathbb{R}, \sigma^2 > 0$.
The posterior distribution of (μ, σ^2) under π_r can be computed as

$$\begin{aligned}
\pi(\mu, \sigma^2) &\propto p(\boldsymbol{x}|\mu, \sigma^2) \times \pi_r(\mu, \sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp(-\tfrac{1}{2} \textstyle\sum (x_i - \mu)^2/\sigma^2) \times (\sigma^2)^{-1} \\
&\propto (\sigma^2)^{-n/2} \exp(-\tfrac{1}{2}[(n-1)s^2 + n(\mu - \bar{x})^2]/\sigma^2) \times (\sigma^2)^{-1} \\
&\big((\sigma^2)^{-(n+1)/2} \exp(-(n-1)s^2/[2\sigma^2])\big) \times \big((\sigma^2)^{-1/2} \exp(-n(\mu - \bar{x})^2/[2\sigma^2])\big),
\end{aligned}$$

where $s^2 = \sum (x_i - \bar{x})^2/(n-1)$. The term on the left is proportional to an inverse-gamma($(n-1)/2, (n-1)s^2/2$) density for σ^2, and the term on the right is proportional to a normal($\bar{x}, \sigma^2/n$) density for μ (conditional on σ^2. This implies that under π_r, the joint distribution of (μ, σ^2) can be described as

$$\begin{aligned} \mu|\bar{x}, \sigma^2 &\sim \text{normal}(\bar{x}, \sigma^2/n) \\ 1/\sigma^2|\bar{x} &\sim \text{gamma}((n-1)/2, (n-1)s^2/2). \end{aligned}$$

The marginal posterior density of $\mu|\bar{x}$ can be obtained by integrating the joint density over σ^2, which gives

$$\frac{\mu - \bar{x}}{s/\sqrt{n}}|X_1, \ldots, X_n \sim t_{n-1}.$$

It is interesting to compare this posterior distribution to that of the sampling distribution of the t-statistic for a fixed value of (μ, σ^2):

$$\frac{\mu - \bar{x}}{s/\sqrt{n}}|\mu, \sigma^2 \sim t_{n-1}.$$

<u>Application (variance estimation)</u>: Consider estimation of σ^2 under the invariant loss

$$L((\mu, \sigma^2), d) = (d - \sigma^2)^2/\sigma^4.$$

The posterior risk is

$$\begin{aligned} \text{E}[(d-\sigma^2)^2/\sigma^4|\boldsymbol{x}] &= \text{E}[(d^2/\sigma^4 - 2d/\sigma^2 + 1|\boldsymbol{x}] \\ &= d^2\text{E}[(\sigma^2)^{-2}] - 2d\text{E}[(\sigma^2)^{-1}] + 1. \end{aligned}$$

Now let $\gamma = 1/\sigma^2$, and let $a = (n-1)/2$ and $b = (n-1)s^2/2$, so that $\gamma|\boldsymbol{x} \sim$ gamma(a, b). The posterior risk in terms of moments of γ is

$$d^2\text{E}[\gamma^2|\boldsymbol{x}] - 2d\text{E}[\gamma|\boldsymbol{x}] + 1$$

which is minimized in d by

$$\begin{aligned}\hat{d}_x &= \frac{\mathrm{E}[\gamma|\boldsymbol{x}]}{\mathrm{E}[\gamma^2|\boldsymbol{x}]} \\ &= \frac{a/b}{a/b^2 + (a/b)^2} \\ &= \frac{ba}{a+a^2} \\ &= \frac{b}{1+a} \\ &= \frac{\frac{1}{2}(n-1)s^2}{\frac{1}{2}(n-1)+1} = \frac{(n-1)s^2}{n+1} = \frac{\sum(x_i-\bar{x})^2}{n+1}.\end{aligned}$$

6.3 Finding the invariant measure

7 Location-scale models Lehmann-style

$X_1,\ldots,X_n \sim$ i.i.d. $P_\theta \in \mathcal{P} = \{p_{01}([x_1-\mu]/\sigma,\ldots,[x_n-\mu]/\sigma)/\sigma^n : \theta = (\mu,\sigma) \in \mathbb{R}\times\mathbb{R}^+\}$

This model is invariant to transformations

$$\mathcal{G} = \{g : \boldsymbol{x} \to a\mathbf{1} + b\boldsymbol{x}, a \in \mathbb{R}, b > 0\},$$

which induce the following group of transformations on Θ :

$$\bar{\mathcal{G}} = \{\bar{g} : (\mu,\sigma^2) \to (a+b\mu, b^2\sigma^2), a \in \mathbb{R}, b > 0\}.$$

Equivariant estimation of σ^r:

Suppose we have an invariant loss under the group $\tilde{\mathcal{G}} = \{\tilde{g} : d \to b^r d, b > 0\}$. An equivariant estimator must satisfy

$$\begin{aligned}\delta(g\boldsymbol{x}) &= \tilde{g}\delta(\boldsymbol{x}) \\ \delta(a\mathbf{1}+b\boldsymbol{x}) &= b^r\delta(\boldsymbol{x}).\end{aligned}$$

In particular, δ must be *invariant* under the additive group:

$$\delta(a\mathbf{1}+\boldsymbol{x}) = \delta(\boldsymbol{x}).$$

We know that any function invariant under a group must be a function of the maximal invariant for that group, which in this case is the differences:

$$\delta(\boldsymbol{x}) = \tilde{\delta}(\boldsymbol{y}(\boldsymbol{x}))$$
$$\boldsymbol{y} = (x_1 - x_n, \ldots, x_{n-1} - x_n).$$

Thus to find the best EE of σ, we can restrict ourselves to EE based on $\boldsymbol{y}$.
What is the joint density of $\boldsymbol{y}$? First find the joint density of $(\boldsymbol{y}, x_n)$, and then integrate over x_n:

$$\begin{aligned} p_{\boldsymbol{y},x_n}(\boldsymbol{y}, x_n|mu, \sigma) &= p_{\boldsymbol{x}}(x_1(\boldsymbol{y}, x_n), \ldots, x_1(\boldsymbol{y}, x_n))|d\boldsymbol{x}/d(\boldsymbol{y}, x_n)| \\ &= p_{01}([y_1 + x_n - \theta]/\sigma, \ldots, [y_{n-1} + x_n - \theta]/\sigma, [x_n - \theta]/\sigma)/\sigma^n \\ p(\boldsymbol{y}|\mu, \sigma) &= \int_{-\infty}^{\infty} p_{01}([y_1 + x_n - \theta]/\sigma, \ldots, [y_{n-1} + x_n - \theta]/\sigma, [x_n - \theta]/\sigma)/\sigma^n dx_n. \end{aligned}$$

Now let $u = (x_n - \theta)/\sigma$, for which $dx = \sigma du$, giving

$$\begin{aligned} p(\boldsymbol{y}|\mu, \sigma) &= \int_{-\infty}^{\infty} p_{01}(y_1/\sigma + u, \ldots, y_{n-1}/\sigma + u, u)/\sigma^{n-1} du \\ &= f(y_1/\sigma, \ldots, y_{n-1}/\sigma)/\sigma^{n-1}. \end{aligned}$$

Thus $\boldsymbol{Y} \sim P_\sigma \in \mathcal{P} = \{f(y_1/\sigma, \ldots, y_{n-1}/\sigma)/\sigma^{n-1}, \sigma > 0\}$.
This family is a scale family, and so the best equivariant estimator based on $\boldsymbol{Y}$ can be obtained from our previous results. In particular:

Characterization: If δ_0 is an equivariant estimator of σ^r, then δ is equivariant iff

$$\delta(\boldsymbol{y}) = \delta_0(\boldsymbol{y})/w(\boldsymbol{z}),$$

where $w(\boldsymbol{z}) = (y_1/y_{n-1}, \ldots, y_{n-2}/y_{n-1}, y_{n-1}/|y_{n-1}|)$.

Optimality: The UMREE estimator is $\delta_0(\boldsymbol{y})/w^*(\boldsymbol{z})$, where $w^*(\boldsymbol{z})$ minimizes

$$\mathrm{E}_{\sigma=1}[L(1, \delta_0(\boldsymbol{y})/w(\boldsymbol{z}))|\boldsymbol{z}].$$

Example (normal variance):
Consider estimation of σ^2 with scaled squared error loss.
The distribution of the y_i's doesn't depend on μ, and
the distribution of the z_i's doesn't depend on μ or σ.
Therefore $\boldsymbol{z}$ is ancillary, and independent of the complete sufficient statistic $(\bar{x}, \sum(x_i - \bar{x})^2)$.

Note that

$$\begin{aligned}(x_i - \bar{x}) &= (x_i - x_n) + (x_n - \bar{x}) \\ &= y_i - \sum y_i/n\end{aligned}$$

for $i = 1, \ldots, n$, with $y_n = 0$. Thus $\sum(x_i - \bar{x})^2$

- is a function of $\boldsymbol{y}$;
- is an equivariant estimator of σ^2.

Taking $\delta_0(\boldsymbol{y}) = \sum(x_i - \bar{x})^2$, the remaining step is to find the optimal $w(\boldsymbol{z})$. Based on our previous results, our best estimator will have the form

$$\delta(\boldsymbol{y}) = \delta_0(\boldsymbol{y})\mathrm{E}_1[\delta_0(\boldsymbol{y})|\boldsymbol{z}]/\mathrm{E}_1[\delta_0^2(\boldsymbol{y})|\boldsymbol{z}].$$

However, recall that $\boldsymbol{y}$ is independent of $\boldsymbol{z}$. Therefore, these conditional expectations are unconditional expectations (constants). We can either calculate them, or note that the estimator must be of the form

$$\delta(\boldsymbol{y}) = c\sum(x_i - \bar{x})^2,$$

and then minimize the risk in c. Taking this latter approach, we find that the UMREE is

$$\delta(\boldsymbol{x}) = \sum(x_i - \bar{x})^2/(n+1).$$

Note that this result says that (under this loss), the MLE and the UMVUE are inadmissible.

Equivariant estimation of μ:

Now let's consider estimation of μ under a loss of the form

$$L((\mu, \sigma^2), d) = \rho((d - \mu)/\sigma).$$

Such a loss is invariant under the group

$$\tilde{\mathcal{G}} = \{\tilde{g} : d \to a + bd, a \in \mathbb{R}^+, b > 0\}.$$

An equivariant estimator must then satisfy

$$\delta(a\mathbf{1} + b\boldsymbol{x}) = a + b\delta(\boldsymbol{x}).$$

In several cases, the best equivariant estimator under the additive group with the scale parameter known

- doesn't depend on the scale;
- is invariant under the linear group.

In these cases, the estimator is also best equivariant of the location problem. To make this more concrete, consider the following strategy:

1. For each fixed σ, find the minimum risk estimator $\delta_\sigma(\boldsymbol{x})$ satisfying

$$\begin{aligned}\delta_\sigma(\boldsymbol{x} + a\mathbf{1}) = \delta_\sigma(\boldsymbol{x}) + a \ \text{ for the model}& \\ \mathcal{P}_\sigma &= \{p([x_1 - \mu]/\sigma, \ldots, [x_n - \mu]/\sigma)/\sigma^n, \mu \in \mathbb{R}\} \\ &= \{p_\sigma(x_1 - \mu, \ldots, x_n - \mu), \mu \in \mathbb{R}\}.\end{aligned}$$

2. If this estimator δ_σ doesn't depend on σ, then
    - it is a valid estimator for the unknown scale problem;
    - it has minimum risk among all estimators equivariant under $\mathcal{G}_a = \{g : \boldsymbol{x} \to \boldsymbol{x} + \mathbf{1}a, a \in \mathbb{R}\}$.

3. If it is also equivariant under the larger group $\mathcal{G} = \{g : \boldsymbol{x} \to a\mathbf{1} + b\boldsymbol{x}, a \in \mathbb{R}^+, b > 0\}$, then it must be UMREE for this problem: Any other estimator equivariant under $\mathcal{G}$ is also equivariant under $\mathcal{G}_a$, and thus must have worse risk than δ_σ.

Example: normal mean

Recall that when σ^2 was known, the UMREE for the additive group was $\delta(\boldsymbol{x}) = \bar{x}$. This is true for each $\sigma^2 > 0$, and so $\bar{x}$ has minimum risk among all estimators satisfying

$$\delta(a\mathbf{1} + \boldsymbol{x}) = a + \delta(\boldsymbol{x}).$$

However, it also satisfies

$$\delta(a\mathbf{1} + b\boldsymbol{x}) = a + b\delta(\boldsymbol{x}),$$

and so it is minimum risk among all such estimators as well.

Nonexample: exponential distribution

Here, $p(x|b,\mu) = 1(x > \mu)e^{(x-\mu)/b}/b, \mu \in \mathbb{R}, b > 0$.

Recall that when b was known, the UMREE for the additive group was $\delta_b(\boldsymbol{x}) = x_{(1)} - b/n$. However, this is not a valid estimator when b is unknown.

8 Invariant testing

Consider two competing location models:

$$\mathcal{P}_0 = \{f_0(x_1 - \mu, \ldots, x_n - \mu) : \mu \in \mathbb{R}\}$$
$$\mathcal{P}_1 = \{f_1(x_1 - \mu, \ldots, x_n - \mu) : \mu \in \mathbb{R}\}.$$

For example f_0 could be the product of standard normal densities, and f_1 could be the product of t-densities (with known degrees of freedom).

Note that each of these models is invariant under the location group $\mathcal{G} = \{g : \boldsymbol{x} \to \boldsymbol{x} + a\mathbf{1}, a \in \mathbb{R}\}$.

We can combine these models into one big model:

$$\mathcal{P} = \{f_k(x_1 - \mu, \ldots, x_n - \mu) : \mu \in \mathbb{R}, k \in \{0, 1\}\}.$$

Let $\theta = \{\mu, k\} \in \mathbb{R} \times \{0, 1\}$. Now since each of $\mathcal{P}_0$ and $\mathcal{P}_1$ are invariant under $\mathcal{G}$, then

$$\boldsymbol{X} \sim P_\theta, \theta = (\mu, 0) \Rightarrow \boldsymbol{X} + a\mathbf{1} \sim P_{\theta'}, \theta' = (\mu + a, 0)$$
$$\boldsymbol{X} \sim P_\theta, \theta = (\mu, 1) \Rightarrow \boldsymbol{X} + a\mathbf{1} \sim P_{\theta'}, \theta' = (\mu + a, 1)$$

Therefore $\mathcal{G}$ induces a group $\bar{\mathcal{G}}$ on $\theta = (\mu, k)$ with elements $\bar{g}$ of the form

$$\bar{g} : (\mu, k) \to (\mu + a, k).$$

Notice

- $\mathcal{P}$ is invariant under $\bar{\mathcal{G}}$
- $\bar{\mathcal{G}}$ is not transitive on $\Theta = \mathbb{R} \times \{0, 1\}$.

Now consider the problem of deciding between $\mathcal{P}_0$ and $\mathcal{P}_1$ with zero-one loss:

$$L(\theta, d) = 1(d \neq k).$$

Is this loss invariant? For it to be invariant, we need to be able to find a class of functions $\tilde{g}$ such that

$$L(\bar{g}\theta, \tilde{g}d) = L(\theta, d)$$
$$L((\mu + a, k), \tilde{g}d) = L((\mu, k), d)$$
$$1(\tilde{g}d \neq k) = 1(d \neq k).$$

This will be satisfied for all a, d, k if $\tilde{g}d = d$. An equivariant estimator in this case is one for which

$$\delta(g\boldsymbol{x}) = \tilde{g}\delta(\boldsymbol{x}) = \delta(\boldsymbol{x}),$$

that is, the equivariant estimators are invariant.

Invariant tests:

You can view a decision rule δ as a test function:

- $\delta : \boldsymbol{x} \to \{0, 1\}$
- $\delta(\boldsymbol{x}) = 0 =$ "say f_0", $\delta(\boldsymbol{x}) = 1 =$ "say f_1"

An equivariant decision rule in this case is an *invariant test*. You can motivate invariant tests as follows: Consider evaluating

$$H_0 : X \sim P_\theta, \theta \in \Theta_0$$
$$H_1 : X \sim P_\theta, \theta \in \Theta_1$$

Suppose both parameter spaces are invariant under $(g, \bar{g})$, so that

$$\bar{g}\Theta_0 = \Theta_0,\ \bar{g}\Theta_1 = \Theta_1.$$

In this case,

- if $X \sim P_\theta, \theta \in \Theta_0$ and $X' = gX$, then $X' \sim P_{\theta'}, \theta' \in \Theta_0$;
- if $X \sim P_\theta, \theta \in \Theta_1$ and $X' = gX$, then $X' \sim P_{\theta'}, \theta' \in \Theta_1$.

The truth of the statement "$\theta \in \Theta_0$" doesn't depend on whether or not your data is X or gX. This suggests test functions $\phi(x) : \mathcal{X} \to [0, 1]$ of the form

$$\phi(gx) = \phi(x).$$

Such tests are called invariant tests.

UMP invariant test

Returning to our location model, we would like to test $\mathcal{P}_0$ versus $\mathcal{P}_1$, i.e.

$$H_0 : \theta \in \mathbb{R} \times \{0\}$$
$$H_1 : \theta \in \mathbb{R} \times \{1\}.$$

This is a "composite versus composite" hypothesis test, but at this point all we know how to test (optimally) are simple versus simple tests. Let's try a simple versus simple version of the above test:

$$\tilde{H}_0;\ \theta = (\mu_0, 0)$$
$$\tilde{H}_1 : \theta = (\mu_1, 1).$$

We can find an optimal invariant test of this simple versus simple test based on the following:

- all invariant tests must be a function of the maximal invariant $\boldsymbol{y}$;
- a MP test of a simple versus simple hypothesis test based on data $\boldsymbol{y}$ can be obtained using the NP lemma.

By the NP lemma, the MP level-α test of $\tilde{H}_0$ versus $\tilde{H}_1$ based on data $\boldsymbol{y}$ has the form

$$\phi(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \frac{p(\boldsymbol{y}|\mu_1,k=1)}{p(\boldsymbol{y}|\mu_0,k=0)} > c \\ 0 & \text{if } \frac{p(\boldsymbol{y}|\mu_1,k=1)}{p(\boldsymbol{y}|\mu_0,k=0)} < c \end{cases}$$

where c is chosen to set the level. To find the specific form of the test, we need to compute the density of $\boldsymbol{y}$ under the two hypotheses. Recall that we have done this calculation before:

$$\begin{aligned} p(\boldsymbol{y}, x_n|\mu_0, k=0) &= p(\boldsymbol{x}_1(\boldsymbol{y}, x_n), \ldots, x_n(\boldsymbol{y}, x_n)|\mu_0, k=0)|\tfrac{d\boldsymbol{x}}{d(\boldsymbol{y},x_n)}| \\ &= f_0(y_1 + x_n - \mu_0, \ldots, y_{n-1} + x_n - \mu_0, x_n - \mu_0) \\ p(\boldsymbol{y}|\mu_0, k=0) &= \int_{-\infty}^{\infty} f_0(y_1 + x_n - \mu_0, \ldots, y_{n-1} + x_n - \mu_0, x_n - \mu_0)\ dx_n \\ &= \int_{-\infty}^{\infty} f_0(y_1 + x_n, \ldots, y_{n-1} + x_n, x_n)\ dx_n \end{aligned}$$

The important thing to note is that this density doesn't depend on the value of μ_0. Similarly,

$$p(\boldsymbol{y}|\mu_1, k=1) = \int_{-\infty}^{\infty} f_1(y_1 + x_n, \ldots, y_{n-1} + x_n, x_n)\ dx_n$$

which doesn't depend on μ_1. The MP test of $\tilde{H}_0$ versus $\tilde{H}_1$ is then to reject $\tilde{H}_0$ when

$$\frac{\int_{-\infty}^{\infty} f_1(y_1 + x_n, \ldots, y_{n-1} + x_n, x_n)\ dx_n}{\int_{-\infty}^{\infty} f_0(y_1 + x_n, \ldots, y_{n-1} + x_n, x_n)\ dx_n} > c,$$

where c is set based on the distribution of $\boldsymbol{y}$ under $\tilde{H}_0$. Since neither the test statistic nor the value of c depend on μ_0 or μ_1, the MP test is the same for all $\mu_0 \in \mathbb{R}$, $\mu_1 \in \mathbb{R}$. Thus the test is the UMP invariant test for the composite versus composite test H_0 versus H_1.

Exercise: Show that the above rejection criterion can be expressed as

$$\frac{\int_{-\infty}^{\infty} f_1(x_1 - \mu, \ldots, x_n - \mu)\ d\mu}{\int_{-\infty}^{\infty} f_0(x_1 - \mu, \ldots, x_n - \mu)\ d\mu} > c,$$

and that the ratio of integrals can be expressed as

$$\frac{p(\boldsymbol{x}|k=1)}{p(\boldsymbol{x}|k=0)} = \frac{\int p(\boldsymbol{x}|\mu, k=1)\pi(\mu)d\mu}{\int p(\boldsymbol{x}|\mu, k=0)\pi(\mu)d\mu}$$

where $\pi(\mu) \propto 1$ is an improper "prior" for μ.

References

James O. Berger. *Statistical decision theory and Bayesian analysis*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1985. ISBN 0-387-96098-8.

Morris L. Eaton. *Group invariance applications in statistics*. NSF-CBMS Regional Conference Series in Probability and Statistics, 1. Institute of Mathematical Statistics, Hayward, CA, 1989. ISBN 0-940600-15-3.

Thomas S. Ferguson. *Mathematical statistics: A decision theoretic approach*. Probability and Mathematical Statistics, Vol. 1. Academic Press, New York, 1967.

E. L. Lehmann and George Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998. ISBN 0-387-98502-6.

E. L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer Texts in Statistics. Springer, New York, third edition, 2005. ISBN 0-387-98864-5.