Admissibility and Minimaxity in Group Families

July 27, 2017

1 The principle of equivariance

Let X be a random observation taking on values in a sample space $\mathcal{X}$ according to a probability distribution from the family

$$\mathcal{P} = \{P_\theta, \theta \in \Omega\}. \tag{1}$$

Let g be a $1:1$ transformations of the sample space onto itself. If for each θ the distribution of $X' = gX$ is again a member of $\mathcal{P}$, say $P_{\theta'}$, and if as θ traverses Ω, so does θ', then the probability model (1) is *invariant* under the transformation g. The transformation g induces a transformation $\bar{g} : \theta \mapsto \theta'$. By definition, $\bar{g}$ is a surjection from Ω onto Ω. Under certain regularity condition, it is also $1:1$.

Now, suppose we would like to estimate $h(\theta)$. Assume $h(\bar{g}\theta)$ depends on θ only through $h(\theta)$, that is,

$$h(\theta_1) = h(\theta_2) \quad \text{implies} \quad h(\bar{g}\theta_1) = h(\bar{g}\theta_2). \tag{2}$$

Let $\mathcal{H} = \{h(\theta) : \theta \in \Omega\}$. Then $\bar{g}$ induces a $1:1$ transformation g^* from $\mathcal{H}$ onto itself.

$$g^* h(\theta) = h(\bar{g}\theta).$$

Definition 1. If the probability model (1) is invariant under g, $h(\theta)$ satisfies (2), and the loss function L satisfies

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

the problem of estimating $h(\theta)$ with loss function L is invariant under g.

Suppose we have decided to use $\delta(X)$ to estimate $h(\theta)$.

- *Functional Equivariance:* It is natural to use

  $$g^* \delta(X) \text{ as the estimator of } g^* h(\theta).$$

- *Formal Invariance:* Consider transformation $X' = gX$. Then $X' \sim P_{\theta'}$ with $\theta' = \bar{g}\theta$. We would like to estimate $g^* h(\theta) = h(\bar{g}\theta) = h(\theta')$. We should use $\delta(X')$ to estimate $h(\theta')$. That is, use

  $$\delta(gX) \text{ to estimate } g^* h(\theta).$$

Instead of a single transformation g, we often consider a group G of transformations. We say a estimation problem is invariant under group G if it is invariant under every $g \in G$.

Definition 2. In an invariant estimation problem, an estimator $\delta(X)$ is said to be *equivariant* if
$$\delta(gX) = g^* \delta(X)$$
for all $g \in G$.

Theorem 1. *If δ is an equivariant estimator in a problem which is invariant under a transformation g, then the risk function of δ satisfies*
$$R(\bar{g}\theta, \delta) = R(\theta, \delta) \text{ for all } \theta.$$

Definition 3. For a group $\bar{G}$ of transformations of Ω, two points $\theta_1, \theta_2 \in \Omega$ are equivalent if there exists a $\bar{g} \in \bar{G}$ such that $\bar{g}\theta_1 = \theta_2$. The totality of points equivalent to a given point (and hence to each other) is called an *orbit* of $\bar{G}$. The group $\bar{G}$ is *transitive* over Ω if it has only one orbit.

For an invariant estimation problem, the risk function of any equivariant estimator is constant on the orbits of $\bar{G}$. If $\bar{G}$ is transitive, the risk of any equivariant estimation is constant. In this case, a minimum risk equivariant (MRE) estimator will typically exist.

2   Equivariant Bayes

Theorem 2. *If the loss function $L(\theta, d)$ is strictly convex in d, a Bayes solution δ_Λ is unique (a.e. $\mathcal{P}$) provided*

- $r(\Lambda, \delta_\Lambda)$ *is finite and*

- *if Q is the marginal distribution of X given by*

$$Q(A) = \int P_\theta(X \in A) d\, \Lambda(\theta),$$

then a.e. Q implies a.e. $\mathcal{P}$.

Proof. Since
$$r(\Lambda, \delta_\Lambda) = \mathrm{E}_X[\mathrm{E}_{\Theta|X}[L(\Theta, \delta_\Lambda(X))]],$$
$\delta_\Lambda(x) = \arg\min_d \mathrm{E}_{\Theta|X=x} L(\Theta, d)$ is a Bayes solution. If there exists another Bayes solution δ'. we have

$$\mathrm{E}_{\Theta|X=x} L(\Theta, \delta'(x)) = \mathrm{E}_{\Theta|X=x} L(\Theta, \delta_\Lambda(x)) \quad \text{a.e. } Q. \tag{3}$$

But $\mathrm{E}_{\Theta|X=x} L(\Theta, d)$ is convex in d, hence (3) implies

$$\delta'(X) = \delta_\Lambda(X) \quad \text{a.e. } Q.$$

$\square$

We shall say that a prior distribution Λ for Θ is *invariant* with respect to $\bar{G}$ if the distribution of $\bar{g}\Theta$ is also Λ for all $\bar{g} \in \bar{G}$. If Λ is invariant, any δ satisfies

$$\mathrm{E}\,R(\Theta, \delta) = \mathrm{E}\,R(\bar{g}\Theta, \delta).$$

But

$$R(\bar{g}\theta, \delta) = \mathrm{E}_{\bar{g}\theta}[L(\bar{g}\theta, \delta(X))] = \mathrm{E}_{\theta}[L(\bar{g}\theta, \delta(gX))] = \mathrm{E}_{\theta}[L(\theta, g^{*-1}\delta(gX))] = R(\theta, g^{*-1}\delta g).$$

Hence

$$\mathrm{E}\,R(\Theta, \delta) = \mathrm{E}\,R(\Theta, g^{*-1}\delta g).$$

If the Bayes estimator δ_Λ is unique, $\delta_\Lambda = g^{*-1}\delta_\Lambda g$ or $\delta_\Lambda(gX) = g^*\delta_\Lambda(X)$, this appears to prove δ_Λ to be equivariant. However, at this point, a technical difficulty arises. Uniqueness can be asserted only up to null sets, that is, sets N with $P_\theta(N) = 0$ for all θ. Moreover, the set N may depend on g. An estimator δ satisfying

$$\delta(gX) = g^*\delta(x) \text{ for all } x \neq N_g$$

where $P_\theta(N_g) = 0$ for all θ is said to be *almost equivariant*.

Theorem 3. *Suppose that an estimation problem is invariant under a group and*

- Λ *is an invariant prior,*

- *the Bayes estimator δ_Λ is unique.*

Then it is almost equivariant.

3 Admissibility and minimaxity in group families

Lemma 1. *A unique Bayes estimator is admissible.*

Proof. If the risk of a unique Bayes estimator is dominated by another estimator, it's Bayes risk also is, which contradicts the uniqueness. $\square$

Theorem 3 implies that there exists an (almost) equivariant estimator which is admissible. Under weak additional assumptions, given any almost equivariant estimator δ, there exists an equivariant estimator δ' which differs from δ only on a fixed null set N.

From now on, we assume $\bar{G}$ is transitive over Ω. Then the risk function of any equivariant estimator is a constant.

Theorem 4. *Under the conditions of Theorem 3, if $\bar{G}$ is transitive over Ω, then δ_Λ is MRE, admissible and minimax.*

Proof. By uniqueness of Bayes estimator, δ_Λ is admissible. An admissible estimator with constant risk is minimax. $\square$

The crucial assumption in this approach is the existence of an invariant prior distribution.

Example 1 (Finite group)**.** Let $X_1, \ldots, X_n$ be iid according to the normal distribution $N(\xi, 1)$. Then, the problem of estimating ξ with squared error loss remains invariant under the two-element group G, which consists of the identity transformation e and the transformation

$$g(x_1, \ldots, x_n) = (-x_1, \ldots, -x_n); \quad \bar{g}\xi = -\xi; \quad g^*d = -d.$$

In the present case, any distribution Λ for ξ which is symmetric with respect to the origin is invariant. If the Bayes estimator is unique, there is a version of the Bayes solution which is equivariant, that is, which satisfies $\delta(-x_1, \ldots, -x_n) = -\delta(x_1, \ldots, x_n)$. The group $\bar{G}$ in this case is, of course, not transitive over Ω.

We shall show in the end of the note that an equivariant Bayes estimator always exists even if Bayes estimator is not unique.

Example 2 (Circular location family)**.** For a number a, let a^* be a number such that $a = 2\kappa\pi + a^*$ and $0 \le a^* < 2\pi$.

Suppose $U_1, \ldots, U_n$ be iid on $(0, 2\pi)$ according to a distribution function F with density f. Suppose that each point is translated on the circle by an amount θ $(0 \le \theta < 2\pi)$ and $X_i = (U_i + \theta)^*$. Since $U_i + \theta \sim f(x_i - \theta)$, we have $X_i \sim f((x_i - \theta)^*)$.

As an illustration. Suppose $Y_1, \ldots, Y_n$ be iid according to $G(y - \eta)$. Let $X_i = Y_i - [Y_i]$, then

$$\Pr(X_i \le x) = \Pr(\sum_{k=-\infty}^{+\infty} \{k \le Y \le k + x\})$$

$$= \sum_{k=-\infty}^{+\infty} (G(x + k - \eta) - G(k - \eta))$$

$$= \sum_{k=-\infty}^{+\infty} (G(x + k - [\eta]) - G(k - [\eta])).$$

Then $2\pi X_i$ is a circular location family.

Consider the problem of estimating θ. The rotation group G, $\bar{G}$ and G^* can be represented by

$$x_i' = (x_i + c)^*, \quad \theta' = (\theta + c)^*, \quad d' = (d + c)^*.$$

If a loss function $L(\theta, d)$ is invariant under G, then

$$L(\theta, d) = L((\theta + c)^*, (d + c)^*).$$

Let $c = -\theta$, then $L(\theta, d) = \rho[(d - \theta)^*]$. Typically, one would want it to depend only on $(d - \theta)^{**} = \min\{(d - \theta)^*, (2\pi - (d - \theta))^*\}$, which is the difference between d and θ along the smaller of the two arcs connecting them. Thus, the loss might

be $((d - \theta)^{**})^2$ or $|(d - \theta)^{**}|$. It is important to notice that neither of these is convex.

An invariant distribution Λ for θ is the uniform distribution over $(0, 2\pi)$. For any Bayes solution δ_Λ, there exists an equivariant estimator δ^* or δ^{**} dominate it (See the end of the note.). Then δ^* or δ^{**} are also Bayes solution and has constant risk. Hence it is admissible ??? and minimax. If the loss function is not convex, δ^{**} can not be used and the equivariant Bayes procedure may be randomized.

Let us next turn to the question of the admissibility and minimaxity of MRE estimators which are Bayes solutions with respect to improper priors.

Example 3 (Location family on the line). Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ has density
$$f(\mathbf{x} - \theta) = f(x_1 - \theta, \ldots, x_n - \theta),$$
and let G and $\bar{G}$ be the groups f translations $x_i' = x_i + a$ and $\theta' = \theta + a$. Lebesgue measure is an invariant measure on Ω.

The posterior density is
$$\frac{f(\mathbf{x} - \theta)}{\int f(\mathbf{x} - \theta)d\theta}.$$
Here we have $\int f(\mathbf{x} - \theta)d\theta < \infty$ a.s. since

$$\int \int f(\mathbf{x} - \theta)d\theta dx_1 \ldots dx_{n-1}$$
$$= \int \int f(x_1 - \theta, \ldots, x_n - \theta)dx_1 \ldots dx_{n-1}d\theta$$
$$= \int \int f(x_1, \ldots, x_{n-1}, x_n - \theta)dx_1 \ldots dx_{n-1}d\theta$$
$$= 1.$$

Therefore, the posterior density is well defined. The generalized Bayes estimator of θ, with loss function is obtained by minimizing the posterior expected loss
$$\frac{\int L(\theta, \delta(\mathbf{x}))f(\mathbf{x} - \theta)d\theta}{\int f(\mathbf{x} - \theta)d\theta}.$$

For the case that L is squared error, the minimizing value of $\delta(x)$ is the posterior expectation of θ, which was seen to be the Pitman estimator.

Since Lebesgue measure is not a probability distribution, Theorem (4) is not applicable and we cannot conclude that the Pitman estimator is admissible or even minimax.

Theorem 5. *Suppose* $\mathbf{X} = (X_1, \ldots, X_n)$ *is distributed according to the density* $f(\mathbf{x} - \theta)$ *and that the Pitman estimator*
$$\delta^*(\mathbf{x}) = \frac{\int \theta f(\mathbf{x} - \theta)d\theta}{\int f(\mathbf{x} - \theta)d\theta}$$
has finite variance. Then, δ^* *is minimax for squared error loss.*

Proof. For any estimator δ,

$$\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta)\Lambda(d\theta) \geq \int R(\theta, \delta_\Lambda)\Lambda(d\theta).$$

Hence a minimax lower bound is

$$\sup_\theta R(\theta, \delta) \geq \sup_\Lambda \int R(\theta, \delta_\Lambda)\Lambda(d\theta).$$

We shall prove that δ^* reaches this lower bound. Let δ_T be the Bayes estimator with respect to the prior

$$\pi_T(u) = \begin{cases} \dfrac{1}{2T} & \text{if } |u| < T \\ 0 & \text{otherwise,} \end{cases}$$

Since δ^* is an equivariant estimator, it has constant risk $r^* = E_0\delta^{*2}(\mathbf{X})$. Since

$$\sup_\Lambda \int R(\theta, \delta_\Lambda)\Lambda(d\theta) \geq \liminf r_T,$$

we only need to show $\liminf r_T \geq r^*$. We begin by establishing the lower bound for r_T

$$r_T \geq (1-\epsilon) \inf_{\substack{a \leq -\epsilon T \\ b \geq \epsilon T}} E_0\, \delta_{a,b}^2(\mathbf{X}), \tag{4}$$

where ϵ is any number between 0 and 1, and $\delta_{a,b}$ is the Bayes estimator with respect to the uniform prior on (a, b) so that, in particular, $\delta_T = \delta_{-T,T}$. We have

$$\delta_{a,b}(\mathbf{x}) = \frac{\int_a^b u f(\mathbf{x} - u)du}{\int_a^b f(\mathbf{x} - u)du}.$$

For any c,

$$\delta_{a,b}(\mathbf{x} + c) = \frac{\int_a^b u f(\mathbf{x} - (u - c))du}{\int_a^b f(\mathbf{x} - (u - c))du}$$

$$= \frac{\int_a^b (u - c) f(\mathbf{x} - (u - c))du}{\int_a^b f(\mathbf{x} - (u - c))du} + c$$

$$= \frac{\int_{a-c}^{b-c} u f(\mathbf{x} - u)du}{\int_{a-c}^{b-c} f(\mathbf{x} - u)du} + c = \delta_{a-c,b-c}(\mathbf{x}) + c.$$

Hence

$$E_\theta(\delta_{-T,T}(\mathbf{X}) - \theta)^2 = E_0[\delta_{-T-\theta,T-\theta}(\mathbf{X})^2].$$

Then

$$r_T = \frac{1}{2T} \int_{-T}^T E_0[\delta_{-T-\theta,T-\theta}(\mathbf{X})^2]d\theta$$

$$\geq (1-\epsilon) \inf_{|\theta| \leq (1-\epsilon)T} E_0[\delta_{-T-\theta,T-\theta}(\mathbf{X})^2]$$

$$\geq (1-\epsilon) \inf_{\substack{a \leq -\epsilon T \\ b \geq \epsilon T}} E_0\, \delta_{a,b}^2(\mathbf{X}).$$

6

Therefore,

$$\liminf_{T \to \infty} r_T \geq (1 - \epsilon) \liminf_{\substack{a \to -\infty \\ b \to +\infty}} E_0 \, \delta_{a,b}^2(\mathbf{X})$$

$$\geq (1 - \epsilon) E_0 \liminf_{\substack{a \to -\infty \\ b \to +\infty}} \delta_{a,b}^2(\mathbf{X})$$

$$= (1 - \epsilon) E_0 \liminf_{\substack{a \to -\infty \\ b \to +\infty}} \frac{\int_a^b u f(\mathbf{x} - u) du}{\int_a^b f(\mathbf{x} - u) du}$$

$$= (1 - \epsilon) E_0 \, \delta^*(\mathbf{X}).$$

$\square$

Thereom 5 is due to Girshick and Savage(1951). The proof given here is due to Peter Bickel.

Theorem 6. *Suppose $X_1, \ldots, X_n$ are iid random variables with density $f(x - \theta)$. If there exists an equivariant estimator δ_0 of θ for which $E_0 \, |\delta_0(\mathbf{X})|^3 < \infty$, then the Pitman estimator δ^* is admissible under squared error loss.*

The admissibility result need not hold when the third-moment condition is dropped.

- For location problem, the MRE estimator has been proved to be admissible under more general loss functions. A key assumption is the uniqueness of the MRE estimator. The MRE estimator is not admissible in the case of nonuniqueness.

- An MRE estimator may be inadmissible in the presence of nuisance parameters, when the corresponding estimator with knwon values of the nuisance parameters is admissible.

- MRE estimators will typically not be admissible except in the simplest situations, but they have a much better chance of being minimax.

- The idea of Theorem 5 is to use a sequence of prior distribution to approximate the Lebesgue measure. We only need to prove the Bayes risk tends to the risk of MRE. To prove the admissibility, we need to consider the convergence rate of the risks.

Example 4 (MRE not minimax). Let the pairs (X_1, X_2) and (Y_1, Y_2) be independent, each with a bivariate normal distribution with mean zero. Let their covariance matrices be $\Sigma = [\sigma_{ij}]$ and $\Delta\Sigma = [\Delta\sigma_{ij}]$, Δ. Consider the problem of estimating Δ with loss function

$$L(\Delta, d) = \begin{cases} 1 & \text{if } \dfrac{|d - \Delta|}{\Delta} > 1/2 \\ 0 & \text{otherwise.} \end{cases}$$

The problem is invariant under the transformation

$$X_1' = a_1 X_1 + a_2 X_2, \quad Y_1' = c(a_1 Y_1 + a_2 Y_2),$$
$$X_2' = b_1 X_1 + b_2 X_2, \quad Y_2' = c(b_1 Y_1 + b_2 Y_2),$$

with $a_1 b_2 \neq a_2 b_1$ and $c > 0$. The only equivariant estimator is $\delta(\mathbf{x}, \mathbf{y}) = 0$. Since $L(\Delta, 0) = 1$, the risk is 1. On the other hand $k^* Y_2^2 / X_2^2$ has risk less than 1 since

$$R = P(|k^* Y_2^2 / X_2^2 - 1| > 1/2) < 1.$$

Example 5 (A raondom walk). Let the parameter space Ω consist of all elements θ of the free group with two generators, that is, the totality of formal products $\pi_1 \ldots \pi_n$ $(n = 0, 1, 2, \ldots)$ where each π_i is one of the elements a, a^{-1}, b, b^{-1} and in which all products $aa^{-1}, a^{-1}a, bb^{-1}, b^{-1}b$ have been canceled. The empty product $(n = 0)$ is denoted by e. The sample point $X = \theta \epsilon$, where ϵ equals one of a, a^{-1}, b, b^{-1} with probability $1/4$. Suppose $\theta = \pi_1 \ldots \pi_n$, then

$$X = \begin{cases} \pi_1 \ldots \pi_n \epsilon & \text{if } \epsilon \neq \pi_n^{-1} \\ \pi_1 \ldots \pi_{n-1} & \text{if } \epsilon = \pi_n^{-1} \end{cases}$$

The problem is to estimate θ based on X. The loss will be 1 if $d \neq \theta$ and 0 otherwise. Suppose $X = \pi_1 \ldots \pi_k$, then a natural estimator of θ is

$$\hat{\theta} = \begin{cases} \pi_1 \ldots \pi_{k-1} & \text{if } X \neq e \\ e & \text{if } X = e \end{cases}$$

Note that as long as ϵ is not the rightmost element of θ, we have $\hat{\theta} = \theta$. Hence $\Pr(\hat{\theta} \neq \theta) \leq 1/4$.

The problem is invariant under transformation

$$X \mapsto \pi_{-r} \ldots \pi_{-1} X, \quad \theta \mapsto \pi_{-r} \ldots \pi_{-1} \theta, \quad d \mapsto \pi_{-r} \ldots \pi_{-1} d.$$

Hence every equivariant estimator $\delta(X)$ must satisfies $g^{-1} \delta(gX) = \delta(X)$. Let $g = X^{-1}$, then $\delta(X) = X \delta(e)$. Then

$$\Pr(\delta(X) \neq \theta) = \Pr(X \delta(e) \neq \theta) = \Pr(\theta \epsilon \delta(e) \neq \theta) = \Pr(\epsilon \delta(e) \neq e) = \Pr(\epsilon \neq e \delta(e)^{-1}) = 3/4.$$

Hence no equivariant estimator is minimax or admissible.

Example 6 (Discrete location family). Let $X = U + \theta$ where U takes on the values $1, 2, \ldots$ with probability

$$P(U = k) = p_k.$$

We observe x and wish to estimate θ with loss function

$$L(\theta, d) = \begin{cases} d - \theta & \text{if } d \geq \theta \\ 0 & \text{if } d \leq \theta. \end{cases}$$

The problem is invariant under arbitrary translation of X, θ and d by the same amount. The only equivariant estimators are those of the form $X - c$. The risk is
$$R(\theta, X - c) = \mathrm{E}_0 \, L(0, X - c) = \sum_{x > c} (x - c) p_x.$$

If $p_x = 1/x(x+1)$, then the risk is infinity. On the other hand, $\delta(x) = x - M|x|$, $M > 1$ has finite risk. In fact,
$$\begin{aligned} R(\theta, X - M|X|) &= \mathrm{E}(X - M|X| - \theta)^+ \\ &= \mathrm{E}(U - M|U + \theta|)^+. \end{aligned}$$

Note that $U - M|U + \theta| > 0$ is equivalent to $\frac{-\theta M}{M+1} < U < \frac{\theta M}{1-M}$, we have
$$\begin{aligned} R(\theta, X - M|X|) &= \mathrm{E}(U - M|U + \theta|)^+ \\ &\leq \mathrm{E} \, U \mathbf{1}_{\{\frac{-\theta M}{M+1} < U < \frac{\theta M}{1-M}\}} \\ &= \sum_{\frac{-\theta M}{M+1} < k < \frac{\theta M}{1-M}} k \frac{1}{k(k+1)} \\ &= \sum_{\frac{-\theta M}{M+1} < k < \frac{\theta M}{1-M}} \frac{1}{(k+1)} \\ &= \int_{\frac{-\theta M}{M+1} - 1}^{\frac{\theta M}{1-M}} \frac{1}{(x+1)} dx \\ &\leq 1 + \int_{\frac{-\theta M}{M+1}}^{\frac{\theta M}{1-M}} \frac{1}{(x+1)} dx \\ &= 1 + \int_{\frac{-M}{M+1}}^{\frac{M}{1-M}} \frac{1}{(x+1)} dx. \end{aligned}$$

3.1 When $\bar{G}$ is not transitive

Let a problem remain invariant relative to the groups G, $\bar{G}$ and G^* over the spaces $\mathcal{X}$, Ω, and D respectively. Then a randomized estimator $\delta(x)$ is defined to be invariant if for all x and g the conditional distribution of $\delta(x)$ given x is the same as that of $g^{*-1}\delta(gx)$.

Suppose $G = \{g_1, \ldots, g_N\}$ is finite. Let $\delta(x)$ be an estimator. Define
$$\delta_i(x) = g_i^{*-1}\delta(g_i x).$$

Consider the randomized estimator δ^* for which
$$\delta^*(x) = \delta_i(x) \text{ with probability } 1/N \text{ for each } i = 1, \ldots, N, \tag{5}$$

and assuming the set D of possible decisions to be convex, the estimator
$$\delta^{**}(x) = \frac{1}{N} \sum_{i=1}^{N} \delta_i(x) \tag{6}$$

9

which, for given x, is the expected value of $\delta^*(x)$.

Note that for a $g \in G$, with probability $1/N$,

$$g^{*-1}\delta^*(gx) = g^{*-1}\delta_i(gx) = g^{*-1}g_i^{*-1}\delta_i(g_igx) = (g_ig)^{*-1}\delta_i(g_igx).$$

As g_i goes through G, g_ig also goes through G. Hence $g^{*-1}\delta^*(gx)$ has the same conditional distribution as $\delta^*(x)$ and it is equivariant. Similarly, g^{**} is also equivariant.

$$R(\theta, \delta^*) = \mathrm{E}_\theta[\mathrm{E}[L(\theta, \delta^*(X))|X]] = \mathrm{E}_\theta[\frac{1}{N}\sum_{i=1}^N L(\theta, \delta_i(X))]$$

$$= \frac{1}{N}\sum_{i=1}^N \mathrm{E}_\theta[L(\theta, g_i^{*-1}\delta(g_iX))] = \frac{1}{N}\sum_{i=1}^N \mathrm{E}_{g_i\theta}[L(\theta, g_i^{*-1}\delta(X))]$$

$$= \frac{1}{N}\sum_{i=1}^N \mathrm{E}_{g_i\theta}[L(\theta, g_i^{*-1}\delta(X))] = \frac{1}{N}\sum_{i=1}^N \mathrm{E}_{g_i\theta}[L(\bar{g}_i\theta, \delta(X))]$$

$$= \frac{1}{N}\sum_{i=1}^N R(\bar{g}_i\theta, \delta) \leq \sup_\theta R(\theta, \delta)$$

If $L(\theta, d)$ is convex in d, $R(\theta, \delta^{**}) \leq R(\theta, \delta^*)$. Hence every minimax estimator can be improved to be equivariant.

Suppose, next, that δ_0 is admissible among al equivariant estimators. If δ_0 is not admissible within the class of all estimators, it is deminated by some δ. But then δ^* dominate δ_0, which is a contradiction.

Suppose G is infinite but there exists an invariant measure K on G with $K(G) < \infty$. Without loss of generality, assume $K(G) = 1$.

Let $\mathbf{g}$ be a random group with distribution K, then δ^* becomes

$$\delta^*(x) = \delta_{\mathbf{g}}(x).$$

and δ^{**} becomes

$$\delta^{**}(x) = \int \delta_g(x)K(dg).$$

Further exploration leads to Hunt-Stein theorem.