*The Annals of Mathematical Statistics*
1971, Vol. 42, No. 4, 1322–1327

# A CHARACTERIZATION OF INVARIANT LOSS FUNCTIONS

BY ROBERT G. STAUDTE, JR.

*Michigan State University*

Maximally invariant loss functions are constructed in a decision theoretic framework, and sufficient conditions for their measurability are given.

**1. Introduction.** This paper is concerned with the implications of group theoretic structure for invariant loss functions. The impetus for this study was a paper by Berk [1] in which equivariant estimators are characterized. While Berk does not discuss optimality in that paper, his results may prove useful in such discussions. The characterization of invariant loss functions presented here will, hopefully, complement his work and further contribute to such discussions.

The invariant estimation problem as defined in [1] is in part generalized to an invariant decision problem in Section 2 so that some examples considered in [1] as well as others may be included. Then in Section 3 "maximally invariant" loss functions are exhibited in this general setting and illustrated by examples; measurability of these loss functions is discussed in Section 4.

**2. Preliminaries.** Our underlying structure consists of a class of probability models $(\mathscr{X}, \mathscr{A}, \mathscr{P})$, a one-one mapping $\psi$ taking $\mathscr{P}$ onto an index set $\Theta$, a measurable space of actions $(\mathscr{Y}, \mathscr{B})$, and a real-valued loss function $L$ defined on $\Theta \times \mathscr{Y}$. We assume that a group $G$ of one-one $\mathscr{A}$-measurable transformations acts on $\mathscr{X}$ and that it leaves the class of models $(\mathscr{X}, \mathscr{A}, \mathscr{P})$ invariant. We further assume that homomorphic images $\bar{G}$ and $\tilde{G}$ of $G$ act on $\Theta$ and $\mathscr{Y}$, respectively. ($\bar{G}$ may be induced on $\Theta$ through $\psi$ as in [1]; and $\tilde{G}$ may be induced on $\mathscr{Y}$ through $L$, see [3].) We shall say that $L$ is invariant if for every $(\theta, y) \in \Theta \times \mathscr{Y}$

$$L(\bar{g}\theta, \tilde{g}y) = L(\theta, y), \qquad g \in G. \tag{2.1}$$

Given the structure described above there are aesthetic and sometimes admissibility grounds for restricting attention to decision rules $\varphi: \mathscr{X} \to \mathscr{Y}$ which are $(G, \tilde{G})$ *equivariant* in the sense that

$$\varphi(gx) = \tilde{g}\varphi(x) \qquad x \in \mathscr{Y}, g \in G. \tag{2.2}$$

If $\tilde{G}$ is trivial and (2) holds, we say $\varphi$ is $G$-invariant, or simply invariant. Further discussion of these concepts may be found in [1], [3], and [4].

**3. Invariant loss functions.** We begin by noting that $L$ is invariant in the sense of (2.1) if and only if $L$ is a $G^{\times}$-invariant function, where $G^{\times}$ is defined on $\Theta \times \mathscr{Y}$ as follows: to each $g \in G$, with homomorphic images $\bar{g}, \tilde{g}$ in $\bar{G}, \tilde{G}$ respectively, let $g^{\times}(\theta, y) = (\bar{g}\theta, \tilde{g}y), (\theta, y) \in \Theta \times \mathscr{Y}$.

Received May 12, 1969; revised January 22, 1971.

Since any invariant loss function will depend on a $G^\times$-maximal invariant, we shall study the latter under various assumptions. A standing assumption throughout the remainder of this paper is that $\tilde{G}$ is a homomorphic image of $\bar{G}$. When context allows, we drop the bars on $\bar{g}, \bar{G}$ for notational convenience.

DEFINITION 3.1. A transformation group $G$ acting on a set $\Theta$ is called *(uniquely) transitive* if for every $\theta$, $\eta \in \Theta$ there exists a (unique) $g \in G$ such that $g\theta = \eta$.

When $G$ is uniquely transitive on $\Theta$ we may index $G$ by $\Theta$: fix an arbitrary point $\theta_0 \in \Theta$ and define $g_\theta$ to be the unique $g \in G$ satisfying $g\theta_0 = \theta$. The identity of $G$ clearly corresponds to $\theta_0$. An immediate consequence is Lemma 3.1.

LEMMA 3.1. *Let $G$ be uniquely transitive on $\Theta$. Fix $\theta_0 \in \Theta$ and define $g_\theta$ as above. Then $g_{h\theta} = hg_\theta$ for $\theta \in \Theta, h \in G$.*

PROOF. The identity $g_{h\theta}\theta_0 = h\theta = hg_\theta\theta_0$ shows that $g_{h\theta}$ and $hg_\theta$ both take $\theta_0$ into $h\theta$, and the lemma follows by unique transitivity.

THEOREM 3.1. *Let $G$ be uniquely transitive on $\Theta$. Fix a reference point $\theta_0 \in \Theta$ and index $G$ by $\Theta$. A maximal invariant $m$ with respect to $G^\times$ acting on $\Theta \times \mathscr{Y}$ is defined by*

$$m(\theta, y) = \widetilde{g_\theta^{-1}}\, y \qquad (\theta, y) \in \Theta \times \mathscr{Y}. \tag{3.1}$$

PROOF. For each $(\theta, y) \in \Theta \times \mathscr{Y}$ and $g \in G$

$$m(g\theta, \tilde{g}y) = (\widetilde{g_{g\theta}^{-1}})\tilde{g}y = (\widetilde{gg_\theta})^{-1}\tilde{g}y = \widetilde{g_\theta^{-1}}\,\widetilde{g^{-1}}\,\tilde{g}y = m(\theta, y)$$

by Lemma 3.1 and the structure preserving properties of homomorphisms. Thus $m$ is $G^\times$-invariant. To see that $m$ is maximal, let

$$m(\theta_1, y_1) = m(\theta_2, y_2).$$

Then $\widetilde{g_{\theta_1}^{-1}}\, y_1 = \widetilde{g_{\theta_2}^{-1}}\, y_2$ or $y_1 = \tilde{g}y_2$ where $\tilde{g} = \tilde{g}_{\theta_1}\widetilde{g_{\theta_2}^{-1}}$.
Since $\theta_1 = g_{\theta_1}\theta_0 = g_{\theta_1}g_{\theta_2}^{-1}\theta_2 = g\theta_2$, $(\theta_1, y_1) = g^\times(\theta_2, y_2)$ for some $g^\times \in G^\times$, and the proof is complete.

In some cases (See Example 1) the transformation group acting on $\Theta$ is not uniquely transitive but it contains a subgroup with this property. By composing certain maximal invariants we may still obtain the $G^\times$-maximal invariant we seek; Theorem 3.2 below provides the details of such a construction. We need one more definition.

DEFINITION 3.2. Let $G$ be a transformation group acting on a space $\Theta$. The *isotropy subgroup of $G$ at $\theta$, $\theta \in \Theta$*, is $G_\theta = \{g \in G : g\theta = \theta\}$.

THEOREM 3.2. *Given the structure defined in Section 2, assume also that $G$ contains a uniquely transitive and normal subgroup $H$. Fix $\theta_0 \in \Theta$ and let $G_0$ be the isotropy subgroup of $G$ at $\theta_0$. If $m_H$ is the $H^\times$-maximal invariant guaranteed by Theorem 3.1 and if $m_0$ is a $\tilde{G}_0$-maximal invariant, then the composition of $m_0$ with $m_H$ is a $G^\times$-maximal invariant.*

PROOF. In order to apply the well-known theorem ([4] page 218) for composing maximal invariants we show first that $H$ and $G_0$ generate $G$, and second that the group induced by $m_H$ and $G_0{}^{\times}$ on the range of $m_H$ is precisely $\tilde{G}_0$.

The subgroups $H$ and $G_0$ of $G$ will generate $G$ if $G \subset HG_0$. Accordingly choose $g$ arbitrarily in $G$ and let $\theta = g\theta_0$. Since also $\theta = h_\theta\theta_0$ we have $h_\theta{}^{-1}g \in G_0$, say $h_\theta{}^{-1}g = g_0$. Hence $g = h_\theta g_0 \in HG_0$ and $G \subset HG_0$. Incidentally, the product $G = HG_0 = G_0H$ is direct if and only if $G_0$ is trivial.

To show that $m_H$ and $G_0{}^{\times}$ induce $\tilde{G}_0$ (in the same way that $s$ and $E$ induce $E^*$ of Theorem 2, page 218, [4]) it suffices to show that $m_H$ is $(G_0{}^{\times}, \tilde{G}_0)$-equivariant; i.e., that

$$m_H(g\theta, \tilde{g}y) = \tilde{g}m_H(y, \theta) \qquad \text{for all } g \in G_0. \tag{3.2}$$

We first show that if $H$ is normal in $G$, then

$$h_{g\theta} = gh_\theta g^{-1} \qquad \text{for all } g \in G_0 \tag{3.3}$$

For any $g \in G_0$, the identity $h_{g\theta}\theta_0 = g\theta = gh_\theta\theta_0$ shows that $h_{g\theta}^{-1}gh_\theta \in G_0$; let $h_{g\theta}^{-1}gh_\theta = g_0$, say. Then $gh_\theta = h_{g\theta}g_0 = g_0h'$ for some $h' \in H$ since $H$ is normal in $G$. It follows that $g_0{}^{-1}g = h'h_\theta{}^{-1} \in H \cap G_0$, which is trivial, so $g = g_0$. Since $g$ was arbitrary in $G_0$ and $g_0 = h^{-1}gh_\theta$, statement (3.3) follows.

The equivalence (3.2) now follows from the identities:

$$m_H(g\theta, \tilde{g}y) = \widetilde{h_{g\theta}^{-1}}\tilde{g}y = \widetilde{gh_\theta{}^{-1}g^{-1}}\tilde{g}y = \tilde{g}(\widetilde{h_\theta{}^{-1}}y) = \tilde{g}m_H(y, \theta).$$

In all the examples to follow, $X$ is a vector of independent random variables which take values in Euclidean $n$-space and $\mathscr{A}$ is the usual Borel field on $\mathscr{X}$.

EXAMPLE 1. Consider the problem of estimating a location parameter for a continuous symmetric distribution. Each component of $X$ has cdf $F_\theta$ defined by $F_\theta(r) = F(r-\theta)$, $r \in R$ for some fixed continuous cdf $F$ which satisfies $F(r)+F(-r) = 1$, $r \in R$. Here $\mathscr{P}$ consists of product measures indexed by the location parameter $\Theta = R = \mathscr{Y}$.

The family of models $\mathscr{P}$ is left invariant under the group $G$ generated by the translation group $H$ and the map $x \to -x$. The reader may easily verify that by Theorems 3.1 and 3.2 any $G^{\times}$ invariant loss function on $\Theta \times$ (range $\varphi$) has the form

$$L(\theta, \varphi(x)) = L(|\varphi(x) - \theta|).$$

EXAMPLE 2. Consider the problem of estimating the location-scale parameter of a distribution belonging to a family generated by a continuous cdf $F$.

$$\mathscr{P} = \left\{P_\theta : F_\theta(x) = F\left(\frac{x-\mu}{\sigma}\right),\ \ x \in R, \theta \in \Theta\right\}$$

$$\Theta = \{(\mu, \sigma) : \mu, \sigma \in R, \sigma > 0\} = \mathscr{Y}.$$

The group $G$ of location and scale changes leaves the class of models invariant.

Since $\bar{G}$ induced on $\Theta$ by $P_\theta \to \theta$ is uniquely transitive, we may apply Theorem 3.1, and obtain invariant loss functions of the form

$$L(\theta, \varphi(x)) = L\left(\frac{\varphi_1(x)-\mu}{\sigma}, \frac{\varphi_2(x)}{\sigma}\right)$$

if $\theta = (\mu, \sigma)$ and $\varphi(x) = (\varphi_1(x), \varphi_2(x))$.

EXAMPLE 3. Let $\Theta$ be the class of all strictly increasing and continuous cdf's, and consider the problem of estimating a numerical characteristic of $F$ when $\mathscr{Y} = R$. This situation arises when, e.g., we estimate a median and our loss depends on both $F$ and the estimate (see (3.4) below).

The family of product measures $\mathscr{P}$ is indexed by $\Theta$ ($\psi(P) = F_p$) and it is left invariant under the group $G$ of strictly increasing $1-1$ onto continuous maps acting coordinatewise on $\mathscr{X}$. The induced $\bar{G}$ acts on $\Theta$ as follows:

$$\bar{g}F = Fg^{-1} = \text{composition of } \quad F \quad \text{with} \quad g^{-1}.$$

We define $\tilde{G}$ to be the group $G$ acting on $\mathscr{Y}$, (with homomorphism $\bar{g} \to g = \tilde{g}$). Now $\bar{G}$ is uniquely transitive on $\Theta$, since for $F_1, F_2 \in \Theta$ the transformation $\bar{g}$ corresponding to $g(x) = F_1^{-1}F_2(x)$ is the unique member of $\bar{G}$ taking $F_1$ into $F_2$. Fix $\theta_0 = F_0$, where $F_0$ is arbitrary in $\Theta$. Then $\bar{g}_F F_0 = F$ implies $g_F = F^{-1}F_0$; by Theorem 3.1 a $G^\times$-maximal invariant is defined by $m(F, y) = \bar{g}_F^{-1}(y) = g_F^{-1}(y) = F_0^{-1}F(y)$. Since $F_0$ is $1-1$ any invariant loss function for this problem has the form

$$L(F, \varphi(x)) = L(F(\varphi(x))).$$

An example of this kind is

$$L(F, \varphi(x)) = |F(\varphi(x)) - \tfrac{1}{2}|. \tag{3.4}$$

EXAMPLE 4. Let $P$, $\Theta$, and $G$ be continued from Example 3 but now expand the action $\mathscr{Y}$ to contain *all* cdf's on $R$. We define $\tilde{G}$ to be $\bar{G}$ acting on $\mathscr{Y}$ as it does on $\Theta$, ($\tilde{g}F = Fg^{-1}$). Then for fixed $F_0 \in \Theta$ we have (as in Example 3) $g_F = F^{-1}F_0$ and $G^\times$-maximal invariant $m(F, \hat{F}) = \tilde{g}_F^{-1}\hat{F} = \hat{F}(F^{-1}(F_0))$ for $(F, \hat{F}) \in \Theta \times \mathscr{Y}$. Since $F_0$ is $1-1$ any invariant loss function has the form $L(F, \hat{F}) = L(\hat{F}(F^{-1}))$. Note that the usual invariant loss functions for this problem—$\sup_x|F(x)-\hat{F}(x)|$ and $\int(F(x)-\hat{F}(x))^2 dF(x)$—are functions of this kind.

**4. Measurability of invariant loss functions.** In the previous sections a loss function was defined to be an arbitrary real-valued function on $\Theta \times \mathscr{Y}$. In order to compute the risk involved in using a decision rule $\varphi$ when $\theta = \psi(P)$ is the parameter associated with the distribution of $X$, we need by definition to find the integral (say, Lebesgue) of $L(\theta, \varphi(X))$ with respect to the measure $P$. Thus it would be desirable to know conditions under which $L$ is measurable in its second argument for fixed $\theta$. If, in addition, we desire to compute Bayes risks with respect to measures on $\Theta$, we want $L$ to be jointly measurable in both arguments. Sufficient conditions for such measurability are presented below for the invariant loss function of Theorem 3.2. Incidentally, we obtain sufficient conditions that the members of $\tilde{G}$ acting on $\mathscr{Y}$ be $\mathscr{B}$-measurable.

Under the conditions for Theorem 3.2, any invariant loss function is given by $L(\theta, y) = L_0(m_0(m_H(\theta, y)))$, $(y, \theta) \in \mathscr{Y} \times \Theta$ where $L_0$: (range $m_0$) $\to R$. If $L_0$ and $m_0$ are measurable (with respect to appropriate $\sigma$-fields) then measurability of $m_H$ in its second argument for fixed $\theta$ implies the same measurability for $L$; a similar statement holds for joint measurability. Therefore we examine the measurability of $m_H$.

Since $m_H(\theta, y) = \widetilde{h_\theta^{-1}}y$, it is clear that $m_H$ is measurable in $y$ for fixed $\theta$ if and only if each $\tilde{h} \in \tilde{H}$ is measurable with respect to $\mathscr{B}$. It is usually assumed that the members of $\tilde{G}$ (and hence $\tilde{H}$) are $\mathscr{B}$-measurable; the following results show that under rather mild conditions such $\mathscr{B}$-measurability is automatically guaranteed.

THEOREM 4.1. *Let $(\mathscr{X},\mathscr{A})$, $(\mathscr{Y},\mathscr{B})$, $G$ and $\tilde{G}$ be given as in Section 2. Assume $\varphi$ is a $(G,\tilde{G})$-equivariant map of $\mathscr{X}$ into $\mathscr{Y}$, and define $\mathscr{B}_\varphi$ to be the collection of $B \subset \mathscr{Y}$ such that $\varphi^{-1}B \in \mathscr{A}$. Then each $\tilde{g} \in \tilde{G}$ is $\mathscr{B}_\varphi$-measurable.*

PROOF. $B \in \mathscr{B}_\varphi \Leftrightarrow \varphi^{-1}B \in \mathscr{A} \Leftrightarrow g^{-1}\varphi^{-1}B \in \mathscr{A} \Leftrightarrow \varphi^{-1}(\tilde{g}^{-1}B) \in \mathscr{A}$ by equivariance $\Leftrightarrow \tilde{g}^{-1}B \in \mathscr{B}_\varphi$.

If we add the natural assumption that $\varphi$ is $(\mathscr{A},\mathscr{B})$-measurable to those stated in Theorem 4.1, then clearly $\mathscr{B} \subset \mathscr{B}_\varphi$. We want $\mathscr{B} = \mathscr{B}_\varphi$ so that we may conclude each $\tilde{g} \in \tilde{G}$ is $\mathscr{B}$-measurable.

COROLLARY 4.1. *If (in addition to the assumptions of Theorem 4.1) $\varphi$ is $(\mathscr{A},\mathscr{B})$-measurable and onto, then either of the following conditions implies $\mathscr{B} = \mathscr{B}_\varphi$:*

(a) *$\varphi(A) \in \mathscr{B}$ for each $A \in \mathscr{A}$*

(b) *$(\mathscr{X},\mathscr{A})$ is a Lusin space and $\mathscr{Y}$ is Euclidean.*

PROOF (a). $B \in \mathscr{B}_\varphi \Leftrightarrow \varphi^{-1}B \in \mathscr{A} \Rightarrow B = \varphi(\varphi^{-1}B) \in \mathscr{B}$. (b) follows directly from a result of Blackwell [2], which is repeated here in our context: Let $(\mathscr{X},\mathscr{A})$ be a Lusin space, let $\varphi$ map $\mathscr{X}$ onto an arbitrary space $\mathscr{Y}$ and define $\mathscr{B}_\varphi$ as above. If $\mathscr{B}$ is a separable subfield of $\mathscr{B}_\varphi$ and every $B \in \mathscr{B}_\varphi$ is a union of atoms of $\mathscr{B}$, then $\mathscr{B} = \mathscr{B}_\varphi$.

To summarize, $m_H$ is measurable in its second argument for fixed $\theta$ if there exists a $(G, \tilde{G})$-equivariant and $(\mathscr{A},\mathscr{B})$-measurable map of $\mathscr{X}$ onto $\mathscr{Y}$ and if $\mathscr{X},\mathscr{Y}$ are "nice" topological spaces; these properties are present in Examples 1, 2, and 3 of the previous section.

In Example 4 the situation is more complicated. The Lévy metric defines a separable Borel field $\mathscr{B}$ on $\mathscr{Y}$ which contains the singletons but Theorem 4.2 is not applicable since an equivariant $\varphi: \mathscr{X} \to \mathscr{Y}$ will not be *onto*; (in fact, it is essentially shown in [4] that if $\varphi$ is equivariant then $\hat{F} = \varphi(x)$ must increase by fixed amounts only on the order statistics). However, Theorem 4.2 does apply when $\mathscr{Y}$ is replaced by the range of $\varphi$ and $\mathscr{B},\mathscr{B}_\varphi$ are replaced by the respective $\sigma$-fields they induce on it. We also observe that if $\tilde{g} \in \tilde{G}$ is $\mathscr{B}_\varphi$-measurable it is measurable with respect to the $\sigma$-field induced by $\mathscr{B}_\varphi$ on the range of $\varphi$. Thus by Theorems 4.1 and 4.2 the loss function of Example 4 is measurable on the range of any $(G,\tilde{G})$-equivariant and $(\mathscr{A},\mathscr{B})$-measurable $\varphi$, and we may meaningfully discuss the integral over $\mathscr{X}$ of $L(F, \varphi(x)) = L_0(\hat{F}(F^{-1}))$ for fixed $F$ and integrable $L_0$. It is not clear that $L$ is measurable over $\mathscr{Y}$ for fixed $F$ since there are uncountably many distinct $\tilde{G}$-orbits which correspond to ranges of different equivariant rules $\varphi$.

Finally, a word about joint measurability of $m_H$ when $\mathscr{Y} = \Theta$ and $\bar{G} = \tilde{G}$. Assume $\mathscr{B}$ is generated by the compact subsets of a topology on $\mathscr{Y}$ and that the group $(\Theta, \cdot)$, where $\theta \cdot \eta = h_\theta \eta$, $(\theta, \eta \in \Theta)$ is topological. Then by definition of topological group $m_H$ is jointly continuous (hence jointly measurable) on $\Theta \times \Theta$. This situation exists in Examples 1 and 2 of the previous section.

**Acknowledgment.** This paper is a revised portion of the author's doctoral dissertation written at the University of Illinois under the guidance of Professor Colin R. Blyth. The author is also indebted to Professor Dennis C. Gilliland of Michigan State University who reviewed the final version of this paper, and to the referee for his helpful comments.

## REFERENCES

[1] BERK, R. H. (1967). A special group structure and equivariant estimation. *Ann. Math. Statist.* **38** 1436–1446.

[2] BLACKWELL, D. (1956). On a class of probability spaces. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **2** 1–6, University of California Press.

[3] FERGUSON, T. S. (1967). *Mathematical Statistics*, Academic Press, New York.

[4] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.