The Annals of Mathematical Statistics
1968, Vol. 39, No. 5, 1573–1576

ON INVARIANCE AND ALMOST INVARIANCE

BY ROBERT H. BERK[1] AND P. J. BICKEL[2]

University of Michigan and University of California, Berkeley

1. Introduction. The requirement that almost-invariant test statistics should be equivalent to invariant test statistics plays a central role in the theory of invariant and unbiased tests [5], sufficiency and invariance [4], etc. A classical condition for this equivalence to hold, due to Stein, may be found in [5], p. 225.

More recently, Bell [1] has given an approach which yields the desired result in nonparametric situations. The purpose of this note is primarily to show that the latter approach applies in most parametric cases as well. In addition, we give a decision-theoretic version of Stein's result.

2. The result. Let $(\mathfrak{X}, \mathcal{B})$ be the measurable (sample) space of the random variable X and $\mathcal{P}$, a family of distributions for X. We suppose $\mathcal{P}$ is generated by G, a group of bimeasurable transformations of $\mathfrak{X}$ to itself, i.e., $\mathcal{P} = \{Pg^{-1} : g \varepsilon G\}$ for any P in $\mathcal{P}$. We refer the reader to [5] for the definitions of terms used from this point on.

Let I be a measurable maximally invariant statistic inducing the invariant σ-field $\mathfrak{I} \subset \mathcal{B}$ and S, another measurable statistic with the induced σ-field $\mathcal{S} \subset \mathcal{B}$, so that the correspondence $X \leftrightarrow (I, S)$ is 1–1 bimeasurable. We suppose also that G acting on X induces a group of transformations, G_S, acting on S. That is, if $X \leftrightarrow (I, S)$, $gX \leftrightarrow (I, g_S S)$. For conditions under which this structure is present, see [3].

(1) THEOREM. *If S is sufficient and boundedly complete, then any $\mathcal{P}$-almost-invariant test function is $\mathcal{P}$-equivalent to an invariant one.*

The proof is preceded by two lemmas. In the sequel, ϕ will denote the test function $\phi(X)$. (ϕ is also called a critical function.) We note that if ϕ is almost-invariant, its distribution is independent of $P \varepsilon \mathcal{P}$. Hence we shall refer to the $\mathcal{P}$-distribution or expectation of almost invariant statistics. Similarly, we may refer to the conditional $\mathcal{P}$-expectation given the sufficient σ-field $\mathcal{S}$ and $E_{\mathcal{P}}(\phi \mid \mathcal{S})$ will denote an element of the $\mathcal{P}$-equivalence class containing $\{E_P(\phi \mid \mathcal{S}) : P \varepsilon \mathcal{P}\}$.

(2) LEMMA. *If ϕ is almost-invariant, $E_P(\phi \mid \mathfrak{I})$ is independent of* of $P \varepsilon \mathcal{P}$.

REMARK. The lemma is actually a special case of the more general fact (which we prove): if $\mathfrak{I}^*$ is the almost invariant σ-field and $\mathcal{C} \subset \mathfrak{I}^*$ is a σ-field, then $E_{\mathcal{P}}(\phi \mid \mathcal{C})$ is meaningful.

PROOF. Choose $P, Q \varepsilon \mathcal{P}$ and let $P = Qg^{-1}$. Then

$$E_P(\phi \mid \mathcal{C}) = E_Q(\phi g \mid g^{-1}\mathcal{C})g^{-1} = E_Q(\phi \mid \mathcal{C})g^{-1} = E_Q(\phi \mid \mathcal{C}) \ [Q].$$

Received 6 February 1968.
[1] Work supported in part by National Science Foundation Grant GP-6008.
[2] Work supported in part by National Science Foundation Grant GP-5059.

The first almost-sure equality is generally true; the second follows because ϕ is almost-invariant and $g^{-1}\mathcal{C} = \mathcal{C}$ $[\mathcal{P}]$; the third, because any $\mathfrak{I}^*$-measurable statistic is almost-invariant. □

(3) Lemma. *$\mathfrak{I}^*$ and $\mathcal{S}$ are independent under every P in $\mathcal{P}$.*

Proof. Since every P in $\mathcal{P}$ induces the same measure on $\mathfrak{I}^*$, this follows from Basu's theorem [5], p. 162, Theorem 2. □

Proof of Theorem. We show that if ϕ is almost-invariant, $\phi = E_{\mathcal{P}}(\phi \mid \mathfrak{I})$ $[\mathcal{P}]$ by establishing that for every B in $\mathcal{B}$ and P in $\mathcal{P}$, $E_P\phi 1_B = E_P(E_{\mathcal{P}}(\phi \mid \mathfrak{I})1_B)$. ($1_B$ denotes the indicator function of B.) Since $\mathcal{B} = \mathfrak{I} \vee \mathcal{S}$, it is sufficient to show this for sets of the form $B = C \cap D$, $C \varepsilon \mathfrak{I}$, $D \varepsilon \mathcal{S}$.

Since $\phi 1_C$ is almost-invariant, by Lemma 2, $E_{\mathcal{P}}(\phi 1_C \mid \mathcal{S}) = E_{\mathcal{P}}\phi 1_C$ $[\mathcal{P}]$. Hence $E_P\phi 1_C 1_D = E_P(1_D E_{\mathcal{P}}(\phi 1_C \mid \mathcal{S})) = E_P 1_D E_{\mathcal{P}}\phi 1_C$. On the other hand, $E_P(E_{\mathcal{P}}(\phi \mid \mathfrak{I})1_C 1_D) = E_P(E_{\mathcal{P}}(\phi 1_C \mid \mathfrak{I})1_D) = E_{\mathcal{P}}\phi 1_C E_P 1_D$, again by Lemma 2. □

3. Applications.

I. *Parametric.* $\mathfrak{X} = R^n$, $\mathcal{B}$ = Borel sets, the coordinates of X are independent with common distribution $P \varepsilon \mathcal{P}$. Theorem 1 applies directly to establish, e.g., that translation invariance and almost-invariance are equivalent under $\mathcal{P}$, the family of normal distributions with unit variance. ($g : (x_1, \cdots, x_n) \to (x_1 + g, \cdots, x_n + g)$, $G \equiv R$; take $I = (X_1 - \bar{X}, \cdots, X_n - \bar{X})$, $S = \bar{X}$.) We obtain a more general result by noting that if $\mathcal{P}^*$ is a set of distributions dominated by $\mathcal{P}$ ($\mathcal{P}^* \ll \mathcal{P}$ if when $B \varepsilon \mathcal{B}$ is $\mathcal{P}$-null, it is also $\mathcal{P}^*$-null), then the $\mathcal{P}$-equivalence of two statistics implies their $\mathcal{P}^*$-equivalence. An obvious criterion that $\mathcal{P}^* \ll \mathcal{P}$ is given by

(4) Lemma. *If for every $P^* \varepsilon \mathcal{P}^*$ there is a $P \varepsilon \mathcal{P}$ so that $P^* \ll P$, then $\mathcal{P}^* \ll \mathcal{P}$.*

The family of normal distributions with unit variance is equivalent to Lebesgue measure. Hence if $\mathcal{P}^*$ is the set of all absolutely continuous distributions on R^n (the coordinates of X need not be either independent or identically distributed), translation invariance and almost-invariance are $\mathcal{P}^*$-equivalent. Similarly, by taking $\mathcal{P}$ = exponential distributions with arbitrary scale, $g : (x_1, \cdots, x_n) \to (gx_1, \cdots, gx_n)$, $G \equiv R - \{0\}$, $S = \sum X_i$, $I = (X_1/S, \cdots, X_n/S)$; we obtain the corresponding result for the scale transformation. Other parametric applications, univariate and multivariate, are easy to provide.

II. *Non-parametric.* (Cf. Bell [1] and Bell and Doksum [2].) Let $\mathfrak{X}$ be the subset of R^n having distinct coordinates and $\mathcal{P}$ be all strictly increasing continuous distributions on R. $(x_1, \cdots, x_n) \to (gx_1, \cdots, gx_n)$, g strictly increasing 1–1 onto R. $S = (X_{(1)}, \cdots, X_{(n)})$, the order statistic, $I = (R_1, \cdots, R_n)$, where R_i is the rank of X_i among $\{X_1, \cdots, X_n\}$. Taking $\mathcal{P}^*$ to be all continuous distributions on R, Lemma 4 applies and $\mathcal{P}^* \ll \mathcal{P}$.

Remark. This result establishes the generality of Loynes theorem [6] dealing with equivariantly distributed estimates of quantiles (see [3] for the terminology). The continuity restriction imposed there on $p_i(\mathbf{X})$ ($\phi(X)$ in our notation) is seen to be unnecessary. (Cf. [6], 500–501.)

We may take $\mathcal{P}^*$ to be the family of all distributions giving the coordinates of

X independent continuous distributions. For if $F_1 \times \cdots \times F_n \varepsilon \mathcal{P}^*$, letting $F = (F_1 + \cdots + F_n)/n$, $F_1 \times \cdots \times F_n \ll F \times \cdots \times F$, hence $\mathcal{P}^* \ll \mathcal{P}$. The independence assumption may also be relaxed; the distribution of X need only be dominated by a product measure (with continuous marginals). A similar result applies to the two sample problem, where the group G acts differently on the first m and remaining n-m coordinates of $\mathfrak{X}$.

4. Stein's Theorem. In this section, we present a decision-theoretic version of Stein's condition, [5] p. 225, for the equivalence of invariance and almost-invariance. In this context, it is actually equivariance that is being considered and we continue with this terminology. Let $(A_0, \mathcal{A})$ be a measurable action space, $\mathcal{A}$ being generated by the countable field $\mathcal{A}_0$. A (randomized) decision rule is a mapping, δ, of $\mathfrak{X}$ into the set of probability measures on $\mathcal{A}$ so that for every A in $\mathcal{A}$, $x \to \delta(x)A$ is $\mathcal{B}$-measurable. Suppose that to each g in G corresponds an $\mathcal{A}$-measurable transformation g^* of A_0 onto itself. Given any δ, we may construct another decision rule, $g\delta$, where $g\delta(x)A = \delta(gx)g^*A$. We say that δ is equivariant if $g\delta = \delta$; it is almost equivariant if $g\delta(X) = \delta(X)$ $[\mathcal{P}]$. (Note that for two decision rules δ and γ, the set

$$[\delta(X) = \gamma(X)] = \cap \{[\delta(X)A = \gamma(X)A] : A \varepsilon \mathcal{A}_0\} \varepsilon \mathcal{B}.)$$

(5) THEOREM. *Suppose there is a measurable structure $\mathcal{G}$ for G and a σ-finite measure ν on $\mathcal{G}$ such that*

(i) $(g, x) \to gx$ *is* $\mathcal{G} \times \mathcal{B} - \mathcal{B}$ *measurable.*

(ii) $(g, a) \to g^{-*}a$ *is* $\mathcal{G} \times \mathcal{A} - \mathcal{A}$ *measurable,* $(g^{-*} = g^{*-1})$.

(iii) $(g, h) \to gh$ *is* $\mathcal{G} \times \mathcal{G} - \mathcal{G}$ *measurable.*

Condition (iii) *permits us to define* νg : *for* $H \varepsilon \mathcal{G}$, $\nu g(H) = \nu(Hg)$.

(iv) *For every* g *in* G, $\nu g \equiv \nu$.

Then if there exists an equivariant decision rule δ_0, every almost-equivariant decision rule is $\mathcal{P}$-equivalent to an equivariant rule.

REMARK. There need not exist equivariant decision rules; see [3]. Without loss of generality, we take ν to be a probability measure.

(6) LEMMA. *For given δ and $D \varepsilon \mathcal{G} \times \mathcal{A}$, with D_g denoting the cross-section of D at g, $\Delta_D : (g, x) \to \delta(x)D_g$ is $\mathcal{G} \times \mathcal{A}$ measurable.*

PROOF. Immediate if D is a measurable rectangle. Also, if $D_i \uparrow (\downarrow) D$, $\Delta_{D_i} \uparrow (\downarrow) \Delta_D$; hence $\{D \varepsilon \mathcal{G} \times \mathcal{A} : \Delta_D$ is $\mathcal{G} \times \mathcal{A}$ measurable$\}$ is a σ-field. □

(7) COROLLARY. *For given δ and $A \varepsilon \mathcal{A}$, conditions* (i) *and* (ii) *imply that* $(g, x) \to g\delta(x)A$ *is* $\mathcal{G} \times \mathcal{B}$ *measurable.*

PROOF. Let $D = \{(g, a) : g^{-*}a \varepsilon A\}$. By (ii), $D \varepsilon \mathcal{G} \times \mathcal{A}$. By (i), $(g, x) \to (g, gx)$ is measurable, hence $(g, x) \to (g, gx) \to \Delta_D(g, gx) = g\delta(x)A$ is measurable. □

PROOF OF THEOREM. Let $N(A) = \{(g, x) : g\delta(x)A \neq \delta(x)A\} \varepsilon \mathcal{G} \times \mathcal{B}$ by Corollary 7. Then $N = \{(g, x) : g\delta(x) \neq \delta(x)\} = \cup\{N(A) : A \varepsilon \mathcal{A}_0\} \varepsilon \mathcal{G} \times \mathcal{B}$. The almost-equivariance condition may be expressed as: for every g in G, the cross-section N_g is $\mathcal{P}$-null. By Fubini, $N_\nu = \{x : \nu N_x > 0\}$ is $\mathcal{P}$-null. Set $\delta_1 =$

$\int_G g\delta\, d\nu$. Then $[\delta_1(X) = \delta(X)] \supset \{x : \delta(x) = g\delta(x)[\nu]\} = \mathfrak{X} - N_\nu$ and δ_1 is equivariant on $\{x : \delta_1(x) = g\delta(x)\ [\nu]\} = \mathfrak{X} - N_G = G(\mathfrak{X} - N_\nu)$, where $GB = \{gx : g \varepsilon G, x \varepsilon B\}$. (Note that for $h \varepsilon G$, $h\delta_1 = \int_G gh\delta\, d\nu = \int_G g\delta\, d\nu h^{-1}$.) Finally, let δ_0 be any equivariant rule and set $\delta^* = \delta_0$ on N_G and $= \delta_1$ on $\mathfrak{X} - N_G$. Then δ^* is equivariant and $\delta(X) = \delta^*(X)\ [\mathcal{P}]$.

In [7], Wesler presents an argument similar to the above. However, his construction involves setting $\delta^* = 0$ on N_G ([7], p. 16 (iii)) which presumably is not an allowable decision rule. He also fails to mention condition (ii), so he appears to have measurability difficulties throughout. (For example, the statement on p. 15, line 3 seems unjustified.) The conditions in the present theorem are slightly more general.

If δ is non-randomized, there is a $\phi : \mathfrak{X} \to A_0$ so that for all x in $\mathfrak{X}$, $\delta(x)\ \{\phi x\} = 1$. If $x \varepsilon \mathfrak{X} - N_\nu$, $\delta_1(x) = \delta(x)$, hence δ_1 is non-randomized on $\mathfrak{X} - N_G$. Choosing δ_0 to be non-randomized yields a non-randomized δ^*. (It is easy to see that the conditions given in [3] for the existence of measurable equivariant rules applies equally for randomized and non-randomized rules.)

REFERENCES

[1] Bell, C. B. (1964). A characterization of multisample distribution-free statistics. *Ann. Math. Statist.* **35** 735–738.

[2] Bell, C. B. and Doksum, K. (1967). Distribution free tests of independence. *Ann. Math. Statist.* **38** 429–447.

[3] Berk, R. H. (1967). A special group structure and equivariant estimation. *Ann. Math. Statist.* **38** 1436–1445.

[4] Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. *Ann. Math. Statist.* **36** 575–614.

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

[6] Loynes, R. M. (1966). Some aspects of the estimation of quantiles. *J. Roy. Statist. Soc. Ser. B.* **28** 497–512.

[7] Wesler, O. (1959). Invariance theory and a modified minimax principle. *Ann. Math Statist.* **30** 1–20.