Statistics 210B, Spring 1998

Class Notes

P.B. Stark

stark@stat.berkeley.edu

www.stat.berkeley.edu/~stark/index.html

January 29, 1998

Second Set of Notes

1 More on Testing and Confidence Sets

See Lehmann, TSH, Ch. 3, 4, 5.

Definition 1 *$S(X)$ is a confidence level $1-\alpha$ confidence set for the parameter $\tau(\mathbf{P}_\theta)$ if*

$$\mathbf{P}_\nu\{S(X) \ni \tau(\nu)\} \geq 1-\alpha \;\; \forall \nu \in \Theta. \tag{1}$$

Definition 2 *Accuracy of Confidence Sets. The accuracy at $\tau(\gamma)$ of the confidence set $S(X)$ for the parameter $\tau(\cdot)$ is*

$$\mathbf{P}_\nu\{S(X) \ni \tau(\mathbf{P}_\gamma)\}, \tau(\nu) \neq \tau(\gamma). \tag{2}$$

That is, it is the probability that the confidence set contains $\tau(\gamma)$, when that is not the true value of τ.

The accuracy is the same as the "false coverage" probability that appeared on the right hand side of the Ghosh-Pratt identity. Typically, when there are nuisance parameters, there is no confidence set $S(X)$ that minimizes 2 for all γ such that $\tau(\nu) \neq \tau(\gamma)$ (there is no *uniformly most accurate* confidence set). However, if one restricts the class of confidence sets in ways that are sometimes reasonable, there are then optimal sets in the restricted class.

Definition 3 *A* $1-\alpha$ *confidence set* $S(X)$ *for the parameter* $\tau(\theta)$ *is* unbiased *if for all* ν,

$$\mathbf{P}_\nu\{S(X) \ni \tau(\gamma)\} \leq 1-\alpha \ \ \forall \gamma \ s.t. \ \tau(\nu) \neq \tau(\gamma). \tag{3}$$

That is, a confidence set is unbiased if the probability of covering a false value of τ is smaller than the confidence level $1-\alpha$. In some sense, a biased confidence set treats some parameter values τ specially. For example, a biased 95% confidence set for the mean θ of a unit variance normal is $\{0\} \cup [X-1.96, X+1.96]$. The coverage probability is 95% for all $\theta \in \mathbf{R}$ except $\theta = 0$, which is in the set with probability one.

The analogous property of a test is that there is no alternative value of τ for which the probability of rejection is less than the level α of the test:

Definition 4 *A level* α *test* δ *of* H *against the alternative* K *is* unbiased *if*

$$\beta_\delta(\mathbf{P}_\nu) \leq \alpha \qquad \forall \nu \in H \tag{4}$$

$$\text{and } \beta_\delta(\mathbf{P}\nu) \geq \alpha \qquad \forall \nu \in K. \tag{5}$$

If the second inequality is strict (>), the test is strictly unbiased.

There exist UMP unbiased tests in many problems for which there is no UMP test, in particular, when $\tau(\nu) \neq \nu$ (when there are nuisance parameters on which the distribution of X depends, but which are irrelevant to the truth of the hypothesis).

Lemma 1 *(Lehmann, TSH, Lemma 4.1.1) Suppose* $\Theta \subset \mathbf{R}$, *and let* Ω *be the common boundary of* $\{\nu : \mathbf{P}_\nu \in H\}$ *and* $\{\nu : \mathbf{P}_\nu \in K\}$. *If* $\mathbf{P}_\Theta$ *is such that the power function* $\beta_\delta(\mathbf{P}_\gamma)$ *is a continuous function of* γ *for every* δ, *and if* δ_0 *is UMP among all level* α *tests such that*

$$\beta_\delta(\gamma) = \alpha \ \ \forall \gamma \in \Omega, \tag{6}$$

then δ_0 *is UMP unbiased.*

This lemma reduces questions about UMP unbiased tests to their behavior on the boundary between the null and alternative, provided the level is a continuous function of the parameter.

In one-parameter exponential families, we saw that there exist UMP tests of one-sided hypotheses about the parameter. For two-sided hypotheses $H : \gamma_1 \leq \theta \leq \gamma_2$ versus $K : \theta < \gamma_1$ or $\theta > \gamma_2$, there exist UMP unbiased tests; the form of their decision functions is (Lehmann, TSH, 4.2)

$$\phi(x) = \begin{cases} 1, & T(x) < c_1,\ T(x) > c_2 \\ a_i & T(x) = c_i,\ i = 1, 2 \\ 0 & c_1 < T(x) < c_2), \end{cases} \tag{7}$$

with c_1 and c_2 chosen s.t.

$$E_{\mathbf{P}_{\gamma_1}} \phi(X) = E_{\mathbf{P}_{\gamma_1}} \phi(X) = \alpha. \tag{8}$$

Example. (Lehmann, TSH, §4.2) Suppose $\mathbf{P}_\Theta$, $\Theta = [0, 1]$ is distributions of the number of successes X in a sequence of a fixed number n of independent trials, each with probability θ of success. (*I.e.*, $X \sim \text{Bin}(n, \theta)$.) This is a one-parameter exponential family. Consider the null hypothesis $H : \mathbf{P}_\theta = \text{Bin}(n, \gamma)$, for fixed $\gamma \in [0, 1]$. The null hypothesis is of the form 7 with $\gamma_1 = \gamma_2 = \gamma$, and $T(x) = x$, so the constraint 8 reduces to

$$E_{\mathbf{P}_\gamma} \phi(X) == 1 - \sum_{x=c_1+1}^{c_2-1} {}_nC_x \gamma^x (1-\gamma)^{n-x} - a_{1n} C_{c_1} \gamma^{c_1} (1-\gamma)^{n-c_1} - a_{2n} C_{c_2} \gamma^{c_2} (1-\gamma)^{n-c_2} = 1 - \alpha \tag{9}$$

Note again that in practice, typically one would choose α so that $a_1 = a_2 = 0$, rather than use a randomized test.

Definition 5 *A family of confidence level* $1 - \alpha$ *confidence sets* $S(X)$ *for* $\tau(\theta)$ *is* uniformly most accurate unbiased *if for all* $\nu \in \Theta$, *it minimizes the probabilities*

$$\mathbf{P}_\nu \{S(X) \ni \tau(\gamma)\} \ \ \forall \gamma \ s.t. \ \tau(\gamma) \neq \tau(\nu). \tag{10}$$

Confidence sets derived by inverting uniformly most powerful unbiased tests are uniformly most accurate unbiased.

2 Equivariant and Invariant Procedures

Definition 6 *A* group *is a set $\mathcal{G}$ and an operation $\circ : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ that satisfies for every g, h, k in $\mathcal{G}$*

1. *$g \circ (h \circ k) = (g \circ h) \circ k$ (associativity)*
2. *There exists a unique $e \in \mathcal{G}$ such that $e \circ g = g \circ e = g$ for every $g \in \mathcal{G}$ (existence of an identity element)*
3. *For each $g \in \mathcal{G}$, there exists $g^{-1} \in \mathcal{G}$ s.t. $g \circ g^{-1} = g^{-1} \circ g = e$.*

Typically, the dot in the notation will be omitted, so we shall write gh (multiplication) in place of $g \circ h$. Another symbol commonly used for the group operation is $+$. Also, while formally a group is the pair $(\mathcal{G}, \circ)$, $\mathcal{G}$ is commonly referred to as the group, with the group operation $\circ$ understood from context.

Definition 7 *A* group of transformations on the set $\mathcal{X}$ *is a collection $\mathcal{G}$ of transformations $g : \mathcal{X} \to \mathcal{X}$ and an operation $\circ : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ such that $(\mathcal{G}, \circ)$ is a group.*

For example, let $\mathcal{X} = \mathbf{R}^n$, and for each $\gamma \in \mathbf{R}^n$, let

$$
\begin{aligned}
g_\gamma : \quad & \mathcal{X} \to \mathcal{X} \\
& \nu \to \nu + \gamma \text{ (vector addition in } \mathbf{R}^n) \qquad (11)
\end{aligned}
$$

Let $+$ denote the group operation, and define $g_\gamma + g_\eta = g_{\gamma+\eta}$. With these definitions, $(\mathcal{G} = \{g_\gamma : \gamma \in \mathbf{R}^n\}, +)$ is a group of transformations on $\mathcal{X}$. The identity element of the group is g_0, the inverse of g_γ is $g_{-\gamma}$, and associativity of the group operation $+$ follows from the associativity of vector addition on $\mathbf{R}^n$. This group is called the *translation group on* $\mathbf{R}^n$.

Definition 8 *Invariance of decision procedures. A statistical decision problem (a set $\mathbf{P}_\Theta$ of distributions on an outcome space $\mathcal{X}$, a loss function L, and a set of possible decisions $\mathcal{D}$) is* invariant *under the group $\mathcal{G}$ of transformations of the outcome space $\mathcal{X}$ if*

1. *The family* $\mathbf{P}_\Theta$ *is closed under* $\mathcal{G}$*, in the sense that for any* $\gamma \in \Theta$ *and any* $g \in \mathcal{G}$*, there exists* $\nu \in \Theta$ *such that if* $X \sim \mathbf{P}_\gamma$*,* $gX \sim \mathbf{P}_\nu$*, with* $\nu \in \Theta$*, and the mapping* $\bar{g} : \Theta \to \Theta$*,* $\gamma \mapsto \nu$ *is one-to-one and onto (*$\bar{g}\Theta = \Theta$*).*

2. *For each* $g \in \mathcal{G}$*, there is a transformation* $h(g) : \mathcal{D} \to \mathcal{D}$ *such that* $h(g_1 g_2) = h(g_1)h(g_2)$*, and* $L(\bar{g}\gamma, h(g)d) = L(\gamma, d)$ *for all* γ*,* g*, and* d*.*

Example. Suppose $\Theta = \mathbf{R}^m$, $\mathbf{P}_\Theta = \{N(\theta, I) : \theta \in \Theta\}$ is the set multivariate normal distributions with independent, unit variance components, $\tau(\theta) = \theta$, $\mathcal{D} = \mathbf{R}^m$, and that $L(\theta, \gamma) = \|\theta - \gamma\|^2$. Let $\mathcal{G}$ be the translation group on $\mathbf{R}^m$. Clearly, the set $\mathbf{P}_\Theta$ is closed under $\mathcal{G}$: if $X \sim \mathbf{P}_\theta$, then $g_\gamma X \sim \mathbf{P}_{\theta+\gamma}$. The set of transformations on $\mathbf{P}_\Theta$ induced by the action of $\mathcal{G}$ on $\mathcal{X}$ is $\bar{\mathcal{G}}$, whose elements $\bar{g}(g_\gamma) = \bar{g}_\gamma$ map $\mathbf{P}_\nu$ to $\mathbf{P}_{\nu+\gamma}$. ($\bar{\mathcal{G}}$ is also a group, with the group operation defined by $\bar{g}_\gamma + \bar{g}_\nu = \bar{g}(g_{\gamma+\nu})$). If we define $h(g_\nu) : \mathcal{D} \to \mathcal{D}$, $\gamma \mapsto \gamma + \nu$, then $h(g_\nu g_\eta)(\gamma) = h(g_\nu)h(g_\eta)\gamma$, and $L(\bar{g}_\nu\theta, h(g_\nu)\gamma) = L(\theta, \gamma)$, as required by the conditions of equivariance.

When the decision problem is invariant under $\mathcal{G}$, it is reasonable to consider only *invariant* decision rules $\delta : \mathcal{X} \to \mathcal{D}$, for which $\delta(gx) = h(g)\delta(x)$. Lehmann (TSH, 1.5) draws a distinction between *invariant* and *equivariant* decision rules: for the former, $h(g)d = d$ for all d, while for the latter, $\delta(gx)$ varies with g. In the invariant case, the decision problem is unchanged under $X \mapsto gX$.

Definition 9 *Suppose we are testing* H *against* K*. Given a transformation* g *on* $\mathcal{X}$*, if* $\{\mathbf{P}_{\bar{g}\nu} : \mathbf{P}_\nu \in H\} = H$ *and* $\{\mathbf{P}_{\bar{g}\nu} : \mathbf{P}_\nu \in K\} = K$*, we say* the problem of testing H against K is invariant under the transformation g.

The set of transformations under which a testing problem is invariant is always a group, with the group operation defined in the natural way; the induced set $\bar{\mathcal{G}}$ of transformations on $\mathbf{P}_\Theta$ also form a group ($\bar{\mathcal{G}}$ is a homomorphism of $\mathcal{G}$).

Definition 10 *A function* $M : \mathcal{X} \to \mathcal{Y}$ *is maximal invariant with respect to the* $\mathcal{G}$ *of transformations on a set* $\mathcal{X}$ *if*

1. $M(x) = M(gx)$ *for all* $g \in \mathcal{G}$ *(*M *is invariant under* $\mathcal{G}$*) and*

2. $M(x) = M(y) \;\Rightarrow\; x = gy$ *for some* $g \in \mathcal{G}$*.*

A test is invariant if and only if it depends on x only through a maximal invariant $M(x)$. For example, suppose the observation X is an iid sample of size n from some common distribution. The distribution of X is clearly invariant under permutations of its components, so we could work instead with the set of order statistics without losing any information about θ. The order statistics are in fact a maximal invariant of the permutation group, so invariant tests regarding θ need depend on X only through its order statistics. Any test that treated the components of X differently would have an *ad hoc* flavor. *Ceteris paribus*, it makes sense to base tests on a maximal invariant function of the data X (with respect to $\mathcal{G}$) if the testing problem is invariant under $\mathcal{G}$.

Subject to some measurability considerations, the set of all invariant tests is characterized by the set of all decision functions $\phi(x) = h(M(x))$ where $M(x)$ is a maximal invariant. Basing tests on sufficient statistics reduces the outcome space $\mathcal{X}$. Invariant tests reduce not only the outcome space $\mathcal{X}$, but also the parameter space Θ:

Theorem 1 *(See Lehmann, TSH, 6.3 Th.3.) If* $M(x)$ *is maximal invariant with respect to* $\mathcal{G}$ *and if* $v(\theta)$ *is maximal invariant with respect to the induced group* $\bar{\mathcal{G}}$*, then the distribution of* $M(X)$ *depends on* θ *only through* $v(\theta)$*.*

The utility of this theorem results from the fact that $M(x)$ and $v(\theta)$ can turn out to be real-valued, even when the dimensions of $\mathcal{X}$ and Θ are large, with the distribution of $M(X)$ having monotone likelihood ratio in $v(\theta)$. That makes it possible to find UMP invariant tests using the one-dimensional optimality theory we saw earlier.

Definition 11 *A test with decision function* ϕ *is* almost invariant with respect to $\mathcal{G}$ *if for all* $g \in \mathcal{G}$*,* $\phi(gx) = \phi(x)$ *except on a* $\mathbf{P}_\Theta$*-null set* $\mathcal{N}_g$ *that can depend on* g*.*

Remark. The power function of a test that is almost invariant under $\mathcal{G}$ is invariant under the induced group $\bar{\mathcal{G}}$. The converse is not true in general.

Remark. Unbiasedness and invariance are not equivalent in general, in that UMP unbiased tests can exist when UMP almost-invariant ones do not, and *vice versa*. However, Lehmann

(6.6, Th. 7) shows that if in a given testing problem, there exists a UMP unbiased test with decision function ϕ^* that is unique up to sets of measure zero, and there also exists a UMP almost-invariant test w.r.t. some group $\mathcal{G}$, then the UMP almost-invariant test is also unique up to sets of measure zero, and the two tests are the same a.e.

Definition 12 *Equivariant Confidence Set. Suppose that the set of distributions* $\mathbf{P}_\Theta$ *on* $\mathcal{X}$ *is preserved under the group* $\mathcal{G}$*, and let* $\bar{\mathcal{G}}$ *be the group of transformations on* Θ *induced by the action of* $\mathcal{G}$ *on* $\mathcal{X}$*. Suppose that the action of* $\bar{\mathcal{G}}$ *on the component* $\tau(\nu)$ *of the more general parameter* ν *depends only on* $\tau(\nu)$*; that is,* $\tau(\bar{g}(\nu)) = \tau(\bar{g}(\gamma))$ *if* $\tau(\nu) = \tau(\gamma)$*. For each* $g \in \mathcal{G}$*, let* $\tilde{g}S = \{\tau(\bar{g}(\nu)) : \tau(\nu) \in S\}$*. If* $S(x)$ *is such that*

$$\tilde{g}S(x) = S(gx) \ \ \forall x \in \mathcal{X}, g \in \mathcal{G}, \tag{12}$$

we say S *is equivariant under* $\mathcal{G}$*.*

Equivariant confidence sets result from inverting invariant tests.