Statistics 210B, Spring 1998

Class Notes

P.B. Stark

stark@stat.berkeley.edu

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

February 5, 1998

Third Set of Notes

1 More on Equivariant Confidence Sets

Definition 1 *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*

$$\tilde{g}S(x) = S(gx) \;\; \forall x \in \mathcal{X}, g \in \mathcal{G}, \tag{1}$$

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

Lehmann (TSH, Ch. 6.11) gives several examples of equivariant confidence sets; the following is taken from there.

Example. Suppose $X = (X_1, X_2)$ has independent, unit variance, normally distributed components with mean $\theta = (\theta_1, \theta_2)$. Let $\mathcal{G}$ be the group of rigid motions of the plane (translations and rotations, but not "distortions"). That is

$$\mathbf{P}_\Theta = \{ \text{ bivariate normal distributions with independent, unit variance components, and mean } \theta \in \mathbf{R}^2\}, \tag{2}$$

and $\Theta = \mathbf{R}^2$. The sample space is $\mathcal{X} = \mathbf{R}^2$ as well. The transformation $\bar{g} \in \bar{\mathcal{G}}$ on $\gamma = (\gamma_1, \gamma_2) \in \Theta$ induced by the action of $g \in \mathcal{G}$ on $x = (x_1, x_2) \in \mathcal{X}$ is just g itself. One equivariant family of confidence sets for θ is

$$S(x) = \{\gamma \in \Theta : (x_1 - \gamma_1)^2 + (x_2 - \gamma_2)^2 \leq c\} \tag{3}$$

(circles centered at the data). To see this, note that for g a rigid-body transformation of the plane, $\gamma \in S(x) \iff \bar{g}\gamma \in S(gx)$. Thus

$$\begin{aligned} \tilde{g}S(x) &= \{\bar{g}\gamma \in \mathbf{R}^2 : \gamma \in S(x)\} \\ &= \{g\gamma : \gamma \in S(x)\} \\ &= \{g\gamma : g\gamma \in S(gx)\} \\ &= \{\nu : \nu \in S(x)\} \\ &= S(x). \end{aligned} \tag{4}$$

Definition 2 *A family of* $1 - \alpha$ *confidence sets for* $\tau(\theta)$ *is* uniformly most accurate equivariant under $\mathcal{G}$ *if it minimizes*

$$\mathbf{P}_\theta\{\tau(\gamma) \in S(X)\} \ \ \forall \gamma \in \Theta \ \ s.t. \ \ \tau(\gamma) \neq \tau(\theta) \tag{5}$$

among all confidence sets $S(x)$ *that are equivariant under* $\mathcal{G}$.

Lemma 1 *Lehmann, TSH, 6.11 Lemma 5. Suppose that for each* $\nu \in \mathbf{T} = \tau(\Theta)$ *there is a group* $\mathcal{G}_\nu$ *of transformations under which the problem of testing* $H : \tau(\theta) = \nu$. *Let* $\mathcal{G}$ *be the group of transformations generated by* $\{\mathcal{G}_\nu\}_{\nu \in \mathbf{T}}$. *Suppose* $S(x)$ *is a* $1 - \alpha$ *confidence procedure that is equivariant w.r.t.* $\mathcal{G}$. *Let* $A(\nu) = \{x : \nu \in S(x)\}$.

1. *The set* $A(\nu)$ *is the acceptance region of a level* α *test of* H, *and it is invariant under* $\mathcal{G}_\nu$ *for each* $\nu \in \mathbf{T}$.

2. *If, for each ν, $A(\nu)$ is a UMP invariant level α test of $H : \tau(\theta) = \nu$, $S(x)$ are confidence level $1 \Leftrightarrow \alpha$ uniformly most accurate equivariant (w.r.t. $\mathcal{G}$) confidence sets.*

Thus if one has a family of tests $A(\nu)$ that are UMP and invariant w.r.t. $\mathcal{G}_\nu$, and $S(x) = \{\nu \in \mathbf{T} : x \in A(\nu)\}$, then $S(x)$ are uniformly most accurate equivariant. That is, uniformly most accurate equivariant confidence sets result from inverting a family of UMP invariant tests.

However, not all problems admit uniformly most accurate equivariant confidence sets, or UMP invariant tests.

Example. (Lehmann, 6.12 Ex. 20.) Suppose $X = (X_1, \cdots, X_n)$ is an i.i.d. sample from a univariate normal distribution, with parameter $\theta = (\mu, \sigma^2) \in \mathbf{R} \times \mathbf{R}^+$ unknown. We want to estimate $\tau(\theta) = \sigma^2$. This problem is invariant under the group $\mathcal{G}$ whose elements translate of all of the components of X by the same constant a. The statistic $S^2 = \sum_j (X_j \Leftrightarrow \bar{X})^2$ is sufficient for σ^2, and is invariant under $\mathcal{G}$. The problem of finding a confidence set for σ^2 is invariant under positive scale changes: $X_j \mapsto bX_j$, $S \mapsto bS$, $\sigma \mapsto b\sigma$, for $b > 0$. If $\sigma^2 \in C(S^2)$ (note change of confidence set to C to keep the traditional definition of S) is an equivariant family of confidence sets, we need $b^2 C(S^2) = C(b^2 S^2)$, which gives

$$\sigma^2 \in C(S^2) \Leftrightarrow \sigma^2/S^2 \in 1/S^2 C(S^2) = C(1). \tag{6}$$

Thus for a confidence set to be equivariant, it must be of the form

$$\sigma^2/S^2 \in C(1), \tag{7}$$

where

$$\mathbf{P}_{(\mu,1)}\{S^{-2} \in C(1)\} = 1 \Leftrightarrow \alpha. \tag{8}$$

This family of confidence sets does not contain one that minimizes the false coverage probability.

Assignment 2. Show that there is no uniformly most accurate confidence set among the collection of confidence sets that satisfy 6 and 8.

This leads one to consider other loss functions for confidence regions (such as the expected volume, which you explored using the Ghosh-Pratt identity in the last assignment).

Lehmann (TSH, 6.12) discusses some optimality measures to use in conjunction with the restriction to equivariant confidence sets, for example, minimizing the measure

$$\int_{C(1)} d\nu \tag{9}$$

(with ν Lebesgue measure on $\mathbf{R}$), or minimizing the scale-invariant measure

$$\int_{C(1)} \nu^{-1} d\nu. \tag{10}$$

This measure has the advantage that if the optimal confidence interval for σ is (σ_-, σ_+), then the optimal confidence interval for σ^r is (σ_-^r, σ_+^r).

Even when there is no group structure to the problem, considering similar measures of the size of a confidence set can lead to practical confidence sets. For example, consider estimating the mean θ of a univariate, unit-variance normal from the observation $X \sim N(\theta, 1)$. We have $\Theta = \mathbf{R}$ and $\mathcal{X} = \mathbf{R}$. Suppose we want to minimize among all confidence *intervals* $S(X)$ the maximum expected length of the interval, whatever be θ. That is, define $R(\theta, S) = E_{\mathbf{P}_\theta}|S(X)|$, where $|S|$ is the diameter of S. The minimax procedure S^* minimizes

$$\sup_{\theta \in \Theta} R(\theta, S) \tag{11}$$

among all $\mathbf{P}_\Theta$-measurable mappings from $\mathcal{X} = \mathbf{R}$ to intervals of $\mathbf{R}$, which we might parametrize by the two (measurable) functions ℓ, u that map x to $S(x) = (\ell(x), u(x))$. Then one way to pick a confidence interval procedure is to minimize

$$\sup_{\theta \in \Theta} R(\theta, S) = \sup_{\theta \in \Theta} E\mathbf{P}_\theta |u(X) \Leftrightarrow \ell(X)|, \tag{12}$$

subject to

$$\inf_{\theta \in \Theta} \mathbf{P}_\theta \{(\ell(X), u(X)) \ni \theta\}. \tag{13}$$

If we let $A(\nu) = (\nu \Leftrightarrow z_{\alpha/2}, \nu + z_{\alpha/2})$, the smallest possible range of observations will be in the acceptance region of each ν, among level α tests. On inverting the tests to get confidence intervals, we would get $\ell(x) = x \Leftrightarrow z_{\alpha/2}$ and $u(x) = x + z_{\alpha/2}$. If we had chosen the acceptance region differently for some ν (for example, picking the region to be an asymmetric interval about ν subject to the level α restriction), the acceptance region would have "reached" to more distant observations, and there would have been some values of x that produced longer confidence intervals. If the set on which we chose asymmetrical intervals had strictly positive measure, this would result in a larger value of the expected length for some value of ν. (For other values, we might have ended up with expected length less than $2z_{\alpha/2}$, but we are interested in the maximum expected length.) *This is not a proof,* but it suggests one.

The "moral," if there is one, is that a quite different consideration from group equivariance or accuracy leads again to the same natural confidence interval. However, this approach through optimization extends quite generally to situations in which there is no group structure, in which Θ is restricted in unusual ways, and in which there is no most powerful test to exploit. The direction this leads is to restrict the class of procedures (confidence sets) through their functional dependence on the data. For example, one might restrict attention to confidence sets that are intervals whose endpoints are affine functionals of the data: $\ell(x) = a + bx$, $u(x) = c + dx$, or even to intervals $\ell(x) = -a + bx$, $u(x) = a + bx$.

In general, the restriction to such procedures can cost a lot, in that the minimax risk over affine procedures might be much larger than the minimax risk over all measurable procedures. However, in some problems, it is possible to bound the "inefficiency" of affine procedures relative to more general nonlinear ones.

Consider, for example, the problem of estimating the mean θ of a unit-variance normal from a single observation $X \sim N(\theta, 1)$, when θ is known to lie in the interval $\Theta = [-c, c]$. This is called the "bounded normal mean" problem. This problem is a cartoon of many inference problems in science in which there are extrinsic physical constraints on the parameter of interest. For example, a spectral absorption coefficient must be between zero and one, and energies must be nonnegative, and can sometimes be bounded above using physical arguments. Donoho (1994, *Ann. Stat.*, *22*, 238-270) shows how to reduce some inference problems about functionals of an infinite-dimensional parameter to questions about the bounded normal mean; we shall visit his work more extensively later.

For the moment, let's consider point estimates that minimize mean-squared error rather, than interval estimates. We seek the estimator $\hat{\theta}^*$ that minimizes

$$\sup_{\theta \in \Theta} E_\theta (\hat{\theta}(X) - \theta)^2 \tag{14}$$

among all $\mathbf{P}_\Theta$-measurable estimators $\hat{\theta}$, where E_θ is shorthand for $E_{\mathbf{P}_\nu}$.

The natural estimator ignoring the constraint $\theta \in \Theta = [-c, c]$ has maximum risk 1 over Θ. The "truncation" estimate

$$\hat{\theta}_t(x) \equiv \begin{cases} x, & |x| \le c \\ c, & x > c \\ -c, & x < -c \end{cases} \tag{15}$$

has maximum risk for $\theta = 0$, for which the risk is

$$\int_{x=-c}^{c} x^2 \phi(x) dx + 2c^2 \Phi(-c), \tag{16}$$

where $\phi(x)$ is the standard normal density and $\Phi(x)$ is the standard normal cdf.

The minimax affine estimate has risk

$$\begin{aligned} R_A &= \min_{a,b} \max_{\theta \in \Theta} E_\theta |aX + b - \theta|^2 \\ &= \min_{a,b} \max_{\theta \in \Theta} E_\theta |aX + b - \theta|^2 \\ &= \min_{a,b} \max_{\theta \in \Theta} E_\theta |a(X - \theta) + (1-a)\theta + b|^2 \\ &= \min_{a,b} \max_{\theta \in \Theta} \{a^2 \mathrm{Var}(X) + ((1-a)\theta + b)^2\}. \end{aligned} \tag{17}$$

We have $\mathrm{Var}(X) = 1$. The risk is quadratic in θ with a positive leading coefficient, so the maximum will be attained at either $-c$ or c:

$$\begin{aligned} \max_{\theta \in [-c,c]} a^2 + ((1-a)\theta + b)^2 &= a^2 + \max\left\{((1-a)c + b)^2, (-(1-a)c + b)^2\right\} \\ &\ge a^2 + ((1-a)c)^2, \end{aligned} \tag{18}$$

so the optimal value of $b = 0$. Stationarity then gives the optimal a as the solution to

$$2a - 2c^2(1-a) = 0 \quad \Rightarrow a = \frac{c^2}{1+c^2}, \tag{19}$$

and thus

$$\begin{aligned} \min_a \max_{\theta \in \Theta} E_\theta |aX - \theta|^2 &= \left[\frac{c^2}{1+c^2}\right]^2 + \left[1 - \frac{c^2}{1+c^2}\right]^2 c^2 \\ &= \left[\frac{c^2}{1+c^2}\right]^2 + \left[\frac{1}{1+c^2}\right]^2 c^2 \\ &= c^2 \frac{1+c^2}{(1+c^2)^2} \\ &= \frac{c^2}{1+c^2}. \end{aligned} \tag{20}$$

Donoho, Liu, and McGibbon (1990, *Ann. Stat., 18*, 1416-1437) show that for this problem the minimax affine risk is no larger than 5/4 of the minimax nonlinear risk.

Let's consider confidence intervals. Suppose we restrict attention to *fixed-length* confidence intervals. That is, we consider intervals of the form

$$\mathcal{I}(x) = \mathcal{I}(\delta, \chi) = [\delta(x) \Leftrightarrow \chi, \delta(x) + \chi], \tag{21}$$

where χ does not depend on the data. To guarantee $1 \Leftrightarrow \alpha$ coverage probability over Θ, we need

$$\inf_{\theta \in \Theta} \mathbf{P}_\theta \{\mathcal{I} \ni \theta\} \geq 1 \Leftrightarrow \alpha. \tag{22}$$

The risk is χ. The minimax risk is

$$\chi_{N,\alpha}(c) = \inf_{\delta \text{ measurable}} \left\{ \chi : \inf_{\theta \in \Theta} \mathbf{P}_\theta \{\mathcal{I} \ni \theta\} \geq 1 \Leftrightarrow \alpha \right\}, \tag{23}$$

and the affine minimax risk is

$$\chi_{A,\alpha}(c) = \min_{a,b} \left\{ \chi : \inf_{\theta \in \Theta} \mathbf{P}_\theta \{[aX + b \Leftrightarrow \chi, aX + b + \chi] \ni \theta\} \geq 1 \Leftrightarrow \alpha \right\}. \tag{24}$$

Because the loss (the half-length of the confidence interval)is not random, the loss is the same as the risk. Clearly $\chi_{N,\alpha}(c) \leq \chi_{A,\alpha}(c)$. Because picking $a = 0$, $b = 0$, allows one to take $\chi = c$, $\chi_{A,\alpha}(c) \leq c$. Furthermore, taking $a = 1$, $b = 0$ allows one to take $\chi = z_{\alpha/2}$, so $\chi_{A,\alpha}(c) \leq z_{\alpha/2}$. Thus

$$\chi_{N,\alpha}(c) \leq \chi_{A,\alpha}(c) \leq \min(c, z_{\alpha/2}). \tag{25}$$

Suppose $c < z_\alpha$. Then clearly, $\chi_{N,\alpha}(c) \leq \chi_{A,\alpha}(c) \leq z_\alpha$, because the interval $[\Leftrightarrow z_\alpha, z_\alpha]$ always covers. However, when the constraint is this restrictive, there is nothing better one can do than pick $[\Leftrightarrow z_\alpha, z_\alpha]$.

To see this, first note that it suffices to consider $\delta(x)$ monotone in x and symmetric about $x = 0$ ($\delta(\Leftrightarrow x) = \Leftrightarrow \delta(x)$). (Why?) Suppose the optimal (nonlinear) $\chi_{N,\alpha} = \chi < c < z_\alpha$. Then there would be no loss in assuming $\delta(x) \leq c \Leftrightarrow \chi$, $x \geq c \Leftrightarrow \chi$, $\delta(x) \geq \Leftrightarrow c + \chi$, $x \leq \Leftrightarrow c$. In order to have coverage probability $1 \Leftrightarrow \alpha$ when $\theta = \Leftrightarrow c$, the interval must be centered at some value of $x < \chi \Leftrightarrow c < 0$ whenever $z_\alpha \Leftrightarrow c > 0$, which is a contradiction, because of the monotonicity and symmetry requirements. The linear rule $\delta(x) = 0$ can be used with $\chi = c$

to get 100% coverage, so $\chi_{N,\alpha}(c) = \chi_{A,\alpha}(c) = c$, $c < z_\alpha$. Furthermore, as the constraint $c \to \infty$, Θ becomes less informative, and $\chi_{A,\alpha}(c)$ and $\chi_{N,\alpha}(c) \to z_{\alpha/2}$; Both the minimax and affine minimax risks are clearly monotone in c. Thus for $c < z_\alpha$, $\chi_{N,\alpha}(c) = \chi_{A,\alpha}(c)$, and for $c \geq z_\alpha$,

$$z_\alpha \leq \chi_{N,\alpha}(c) \leq \chi_{A,\alpha}(c) \leq z_{\alpha/2}, \tag{26}$$

which implies

$$\frac{\chi_{A,\alpha}(c)}{\chi_{N,\alpha}(c)} \leq \frac{z_{\alpha/2}}{z_\alpha}. \tag{27}$$

For $\alpha = 0.05$, this ratio is about $1.96/1.645 = 1.19$. Thus in this problem, the minimax affine fixed-length 95% confidence interval is at most about 20% longer than the minimax fixed-length 95% confidence interval.

Problem. (not assigned) Show that it indeed suffices to consider monotone, symmetric rules $\delta(x)$ at which to center a fixed-length confidence interval for the bounded normal mean.

It is a rather unsatisfactory property of the minimax fixed-length interval in this problem that if the prior information is sufficiently strong (*e.g., if* $c < z_\alpha$*), the optimal procedure ignores the data and just returns the prior information. This is an artifact of looking only at the worst-case behavior.*

In contrast, one might consider intervals whose lengths depend on the data, for example, the truncation interval

$$\mathcal{I}_t(x) = [x \Leftrightarrow z_{\alpha/2}, x + z_{\alpha/2}] \cap [\Leftrightarrow c, c]. \tag{28}$$

This interval has random length, but the half-length never exceeds $\min\{c, z_{\alpha/2}$. *A different criterion of optimality one might consider (rather than the fixed-length) is*

$$R^+(\mathcal{I}) \equiv \sup_{\theta \in \Theta} E_\theta |\mathcal{I}(X)|. \tag{29}$$

This is the largest expected length of the confidence interval.

Assignment 3. *Find* $R^+(\mathcal{I}_t)$ *for* $\alpha = 0.05$, $c = 1/2, 1, 2, 5, 10$. *Compare with* $R^+($*affine minimax fixed-lengt* *Note that the risk of the affine minimax procedure is just its fixed length, whether measured by* R^+ *or by the loss function for which it was derived.* Hint. *For what value of* $\theta \in \Theta$ *is the maximum expected length attained?*

2 Bayesian Credible Regions

The Bayesian analogue of a confidence set is a credible region*. To construct a credible region, one must think of the parameter $\tau(\theta)$ as itself being a random variable. Denote its (prior) distribution, which is assumed to be known completely, by* $\Pr$.

Definition 3 *A level $1 \Leftrightarrow \alpha$ credible region for the parameter $\tau(\theta)$ is a set $S(x)$ such that*

$$\Pr\{\tau(\theta) \in S(x) | X = x\} \geq 1 \Leftrightarrow \alpha. \tag{30}$$

That is, S is a set such that the posterior probability that S contains τ, given the data, is at least $1 \Leftrightarrow \alpha$.

Note that there is not a unique S with this property; a criterion often used to obtain a unique S is to take S to be a level set of the posterior distribution of θ. Another is to introduce a loss function associated with a measure of the "size" or "volume" of the confidence set (as we have been discussing in a frequentist context), and to find the region that minimizes that loss (or the risk) subject to the posterior coverage constraint.

Assignment 4: Bounded Normal Mean *Suppose $\mathbf{P}_\Theta = \{N(\gamma, 1) : \gamma \in [\Leftrightarrow a, a]\}$, $X \sim N(\theta, 1)$, $\Theta = [\Leftrightarrow a, a]$, $\theta \sim U[\Leftrightarrow a, a]$. Let $\mathcal{D}$ be the set of Lebesgue-measurable subsets of* $\mathbf{R}$. *Let $L(\theta, d) = \mu(d)$, where $\mu(d)$ is the Lebesgue measure of the set d. (1) Characterize (give equations that determine it) a $1 \Leftrightarrow \alpha$ credible region d for θ that minimizes L given $X = x$. (2) Is it sufficient to assume that d is an interval? (3) Does this approach also minimize the risk $E_{\Pr} L = E_{\Pr} \mu(d)$? (4) Find the credible region explicitly for $a = 1$, $x = 0, 0.5, 1, 2, 10$. (5) For $a = 1$, find explicitly or estimate by simulation the risk $E_{\Pr} L = E_{\Pr} \mu(d)$ of this procedure.*