Theory of Statistical Inference - Lecture IV.5
STA422 and STA2162

Michael Evans
University of Toronto
https://utstat.utoronto.ca/mikevans/sta422/sta4222026.html

2026

- unbiasedness will sometimes work to obtain an optimal unbiased decision procedure although even in that case we showed that an optimal unbiased procedure is not admissible

- unbiasedness seems applicable for some parameters when $\{f_\theta : \theta \in \Theta\}$ is an exponential family

- the other class of models that have a specific structure are the group families and for these contexts invariance is the applicable restriction principle

- to define a group model you need the definition of a group

Definition IV.5.1 A group is a set G with a product $\cdot$ defined on it such that if $g_1, g_2 \in G$ then $g_1 \cdot g_2 \in G$ that satisfies (i) (associative) $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$ (identity) there is an element $e \in G$ such that $e \cdot g = g \cdot e$ (iii) (inverse) for each $g \in G$ there exists $g^{-1} \in G$ s.t. $g^{-1} \cdot g = g \cdot g^{-1} = e$. ■

Example IV.5.1

- additive group $(G, \cdot) = (\mathbb{R}^k, +)$ ($g \in \mathbb{R}^k$, $g_1 \cdot g_2 = g_1 + g_2, e = 0, g^{-1} = -g$)

- mutiplicative group $(G, \cdot) = ((0, \infty), \times)$ ($g =$ a nonnegative real number, $g_1 \cdot g_2 = g_1 g_2, e = 1, g^{-1} = 1/g$)

- full matrix group $(G, \circ) = (\{g \in \mathbb{R}^{k \times k} : A$ is invertible$\}, \cdot =$ matrix product$)$ $(e = I)$

- also for matrix group could take $G =$ set of all $k \times k$ lower (upper) $\triangle$-matrices with positive diagonal elements or $G =$ set of all $k \times k$ orthogonal matrices ■

Definition IV.5.2 Group G *acts* on a set $\mathcal{X}$ when $g : \mathcal{X} \to \mathcal{X}$ and (i) $e(x) = x$, (ii) $g_1 \cdot g_2(x) = g_1(g_2(x))$. ■

- note - G can now be considered as a group of transformations on $\mathcal{X}$ with the product being composition $\circ$

- this expresses the idea that some object possesses certain symmetries - e.g. $\mathcal{X} =$ circle in $\mathbb{R}^2$ and $G =$ set of rotations about the origin as these leave the circle invariant (it remains the same set after transformation)

- so suppose $\mathcal{X}$ = sample space for $\{f_\theta : \theta \in \Theta\}$ and G acts on $\mathcal{X}$

- when $x \sim f_\theta$, then $y = g(x)$ has density (change of variable)

$$f_\theta(g^{-1}(y))J_g(g^{-1}(y))$$

where $J_g(x) = |\det(\partial g(x)/\partial x)|^{-1/2}$ and $\partial g(x)/\partial x$ is the Jacobian matrix of g

Definition IV.5.3 The model $\{f_\theta : \theta \in \Theta\}$ is *invariant* under the transformation group G acting on $\mathcal{X}$ whenever, for all $\theta \in \Theta, g \in G$

$$f_\theta(g^{-1}(y))J_g(g^{-1}(y)) \in \{f_\theta : \theta \in \Theta\}. \blacksquare$$

- a bit simpler to work with the probability measures P_θ

- note - underlying prob. measure P on Ω and $X : \Omega \to \mathcal{X}$ then $P_X(B) = P(X \in B) = P(X^{-1}B)$ and we write $P_X(B) = P \circ X^{-1}(B)$

- so another way to express the invariance of the model under the action of G is to say that $P_\theta \circ g^{-1} \in \{P_\theta : \theta \in \Theta\}$ for every $\theta \in \Theta, g \in G$

- for $g \in G$, define $\bar{g} : \Theta \longrightarrow \Theta$ by $P_{\bar{g}(\theta)} = P_\theta \circ g^{-1}$

Lemma IV.5.1 G acts on Θ via $-$.

Proof: For $\overline{g_1 \cdot g_2}$ we have

$$\begin{aligned} P_{\overline{g_1 \cdot g_2}(\theta)}(B) &= P_\theta((g_1 \circ g_2)^{-1} B) = P_\theta(g_2^{-1} \circ g_1{}^{-1} B) \\ &= P_{\bar{g}_2(\theta)}(g_1^{-1} B) = P_{\bar{g}_1 \circ \bar{g}_2(\theta)}(B) \end{aligned}$$

and so $\overline{g_1 \cdot g_2} = \bar{g}_1 \circ \bar{g}_2$ and $\bar{e}$ is the identity since $P_\theta \circ e^{-1} = P_\theta$. ■

- note for a sample $(x_1, \ldots, x_n)$ the group $(G, \cdot)$ acts as $g(x_1, \ldots, x_n) = (g(x_1), \ldots, g(x_n))$

Definition IV.5.3 The model $\{f_\theta : \theta \in \Theta\}$ is a *group model* if for some transformation group G acting on $\mathcal{X}$ and any θ_0 there is unique $g \in G$ s.t. $P_\theta = P_{\bar{g}(\theta_0)}$.

Example IV.5.2 *location normal*

- $x_1, \ldots, x_n$ iid $N_k(\mu, \Sigma_0)$ where $\mu \in \mathbb{R}^k$ is unknown and Σ_0 is known

- here $(G, \cdot) = (\mathbb{R}^k, +)$ and when $x \sim N_k(\mu, \Sigma_0)$, then $g(x) = x + g \sim N_k(g + \mu, \Sigma_0)$ so the group leaves the model invariant

- also, for any μ_0 and $z \sim N_k(\mu_0, \Sigma_0)$ we have $x = g(z) \sim N_k(\mu, \Sigma_0)$ when $g = \mu - \mu_0$ and so this is a group model, typically take $\mu_0 = 0$

- also $\bar{g}(\mu) = g + \mu$ defined on $\mathbb{R}^k$ ■

- what does it mean for a group $(G, \cdot)$ to leave the decision problem $I^{dec} = (\{f_\theta : \theta \in \Theta\}, Loss, x)$ invariant?

- the group needs to leave the model invariant (as described) and the loss *Loss* invariant

- recall decision goal is to estimate $\psi = \Psi(\theta)$

- now $\psi_{true} = \Psi(\theta_{true})$ so if we transform the data x to $g(x)$ the true value of θ is now $\bar{g}(\theta_{true})$ and so we want $\Psi(\bar{g}(\theta_{true}))$ to equal the new true value

- for this we require Ψ be *equivariant*, namely, $\Psi(\bar{g}(\theta_1)) = \Psi(\bar{g}(\theta_2))$ iff $\Psi(\theta_1) = \Psi(\theta_2)$ and then we can define an action of G on $\Psi(\Theta)$ by $g_*(\psi) = \Psi(\bar{g}(\theta))$ where $\psi = \Psi(\theta)$

Lemma IV.5.2 When Ψ is equivariant, then G acts on $\Psi(\Theta)$ via $*$.
Proof: We have

$$\begin{aligned}(g_1 \cdot g_2)_*(\psi) &= \Psi(\overline{g_1 \cdot g_2}(\theta)) = \Psi(\bar{g}_1 \circ \bar{g}_2(\theta)) \\ &= g_{1*}(\Psi(\bar{g}_2(\theta)) = g_{1*} \circ g_{2*}(\Psi(\theta))\end{aligned}$$

so $(g_1 \cdot g_2)_* = g_{1*} \circ g_{2*}$, e_* is the identity. ■

Example IV.5.2 *location normal (continued)*

- the parameter μ is equivariant because $g + \mu_1 = g + \mu_2$ iff $\mu_1 = \mu_2$ but $\psi = \Psi(\mu) = \mu^t \mu$ is not equivariant since we can have $\mu_1^t \mu_1 = \mu_2^t \mu_2$, but in general

$$(\mu_1 + g)^t (\mu_1 + g) \neq (\mu_2 + g)^t (\mu_2 + g),$$

e.g. $\mu_1^t \mu_1 = \mu_2^t \mu_2 = 1$ but $\mu_1^t g \neq \mu_2^t g$ so not all parameters are amenable to this structure ■

Definition IV.5.4 The decision problem $I^{dec} = (\{f_\theta : \theta \in \Theta\}, Loss, x)$ is invariant under the group G if the model and *Loss* are invariant under the actions G where the invariance of the loss function means

$$Loss(\theta, \psi) = Loss(\bar{g}(\theta), g_*(\psi))$$

for every $\theta \in \Theta, \psi \in \Psi(\Theta), g \in G$. ■

Example IV.5.2 *location normal (continued)*

- suppose quadratic loss is used then

$$\begin{aligned} Loss(\bar{g}(\mu), g_*(\mu')) &= Loss(g+\mu, g+\mu') \\ &= (g+\mu-g-\mu')^t A(\cdot) = Loss(\mu,\mu') \end{aligned}$$

and so the location group leaves the problem invariant. ■

Lemma IV.5.3 The group G acts on $\mathcal{D}(I^{dec})$ via $\tilde{g}(\delta)$ given by $\tilde{g}(\delta)(x, A) = \delta(g^{-1}(x), g_*^{-1}A)$.
Proof: We have

$$\begin{aligned} \widetilde{g_1 \cdot g_2}(\delta)(x, A) &= \delta(g_2^{-1} \circ g_1^{-1}(x), g_{2*}^{-1} \circ g_{1*}^{-1}A) = \widetilde{g}_2(\delta)(g_1^{-1}(x), g_{1*}^{-1}A) \\ &= \widetilde{g}_1 \circ \widetilde{g}_2(\delta)(x, A) \end{aligned}$$

so $\widetilde{g_1 \cdot g_2} = \widetilde{g}_1 \circ \widetilde{g}_2$ and $\widetilde{e}(\delta)(x, A) = \delta(e^{-1}(x), e_*^{-1}A) = \delta(x, A)$. ■

Definition IV.5.5 When the group G leaves the decision problem $\mathcal{D}(I^{dec})$ then a decision function δ is invariant if $\tilde{g}(\delta) = \delta$ for every $g \in G$ let $\mathcal{D}_G(I^{dec})$ denote the class of all invariant (under the action of G) decision functions. ■

- the invariance principle in decision theory is then to find the optimal $\delta \in \mathcal{D}_G(I^{dec})$ (if it exists)

Lemma IV.5.4 When the group G leaves the decision problem $\mathcal{D}(I^{dec})$ invariant, a nonrandomized decision function δ (or d) is invariant iff d is equivariant, i.e., $d(g(x_1)) = d(g(x_2))$ iff $d(x_1) = d(x_2)$ so $d(g(x)) = g_*(d(x))$.
Proof: We have $\delta(x, A) = 1$ iff $d(x) \in A$ and is 0 otherwise. Now $\delta = \tilde{g}(\delta)$ implies

$$\begin{aligned} \tilde{g}(\delta)(x, A) &= \delta(g^{-1}(x), g_*^{-1}A) = \begin{cases} 1 & d(g^{-1}(x)) \in g_*^{-1}A \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & g_* d(g^{-1}(x)) \in A \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

iff (taking $A = \{d(x)\}$), $g_* d(g^{-1}(x)) = d(x)$. ■

- so in a convex decision problem we search for the optimal d which, when it exists is called a *Pitman estimator*

Lemma IV.5.5 When the group G leaves the decision problem $\mathcal{D}(I^{dec})$ invariant, then $R(\bar{g}(\theta), \widetilde{g}(\delta)) = R(\theta, \delta)$ for all $\theta \in \Theta, g \in G$.
Proof: We have

$$\begin{aligned}
R(\bar{g}(\theta), \widetilde{g}(\delta)) &= \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\bar{g}(\theta), \psi) \widetilde{g}(\delta)(x, d\psi) P_{\bar{g}(\theta)}(dx) \\
&= \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\bar{g}(\theta), g_* \circ g_*^{-1}(\psi)) \delta \circ g_*^{-1}(g^{-1}(x), d\psi) P_{\bar{g}(\theta)} \\
&= \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\bar{g}(\theta), g_*(\psi)) \delta(g^{-1}(x), d\psi) P_{\bar{g}(\theta)}(dx) \\
&= \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\theta, \psi) \delta(g^{-1}(x), d\psi) P_\theta \circ g^{-1}(dx) \\
&= \int_{\mathcal{X}} \int_{\Psi(\Theta)} L(\theta, \psi) \delta(x, d\psi) P_\theta(dx) = R(\theta, \delta). \blacksquare
\end{aligned}$$

Corollary IV.5.1 If δ is invariant, then $R(\bar{g}(\theta), \delta) = R(\theta, \delta)$ for all $\theta \in \Theta, g \in G$ and, if the group acts transitively on Θ (for any θ_1, θ_2 there exists g s.t. $g(\theta_1) = \theta_2$) then $R(\theta, \delta)$ is constant in θ.

Example IV.5.2 *location normal (continued)*

- with quadratic loss this problem is convex so we can restrict to nonrandomized estimators and we can reduce to dependence on the data only through the mss $\bar{x}$

- so an estimator is of the form $d(\bar{x})$ and to be equivariant we must have $d(\bar{x}) = \bar{x} + d(0)$ since $\bar{x} \in G$ and the optimal invariant estimator is given by the constant $d(0)$ that minimizes

$$\begin{aligned} MSE(\mu, d) &= E_\mu((d(\bar{x}) - \mu)^t A(d(\bar{x}) - \mu)) \\ &= (\bar{x} + d(0) - \mu)^t A(\bar{x} + d(0) - \mu) \\ &= E_\mu((\bar{x} - \mu)^t A(\bar{x} - \mu)) + d^t(0) A d(0) \\ &= E_\mu((\bar{x} - \mu)^t A(\bar{x} - \mu)) + d^t(0) A d(0) \end{aligned}$$

- since $d^t(0)Ad(0) \geq 0$ and equal to 0 only when $d(0) =$, the optimal invariant estimator is $d(\bar{x}) = \bar{x}$ ■

Personal Opinion

Unbiasedness and invariance both lead to some interesting results, but neither provides a theory that is generally applicable to all decision problems. Both do not lead to admissible procedures even when they work.

More relevant are the invariance principles presented when discussing Birnbaum's Theorem, namely, invariance on the parameter space and invariance on the sample space. These said that 1-1 (smooth) transformations of the data or the parameter space should not lead to inferences that are essentially different. For example, if we estimate θ *by* $d(x)$ *and we reparameterize to* $\psi = \Psi(\theta)$, *via 1-1* Ψ, *and transform the data to* $t = T(x)$, *via 1-1* T, *then any valid theory of statistical inference should estimate* ψ *by* $\Psi(d(T^{-1}(t)))$. *Unbiasedness and invariance do not lead to inferences with this more general invariance property.*