Theory of Statistical Inference - Lecture V
STA422 and STA2162

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

2026

V Fiducial Inference and Structural Inference

- this is an idea due to Fisher who probably recognized a problem with the confidence concept, namely, if we report a γ-confidence region $C_\Psi(x)$ for $\psi = \Psi(\theta)$, so

$$P_\theta(C(x)) \geq \gamma$$

for every $\theta \in \Theta$, then we would like to say "the probability that the true value of ψ is in $C_\Psi(x)$ is at least γ" but we can't

- instead we say "if we repeated this experiment many (infinite) times, a proportion at least γ would result in the true value being in the region" which doesn't seem that compelling when considering the given data

- this together with the other counter-intuitive behavior of the confidence concept leads to a search for a better approach and Fisher's fiducial is one such attempt

Definition V.1 For model $\{f_\theta : \theta \in \Theta\}$ suppose that $T : \Theta \times \mathcal{X} \rightarrow \mathbb{R}$ is pivotal ($T(\theta, x)$ has the same distribution for every θ). Then for $A \subset \Theta$ the fiducial probability of A, when x has been observed, is given by

$$P_{fiducial,x}(A) = P(\{t : T(\theta, x) = t \text{ for some } \theta \in A\}). \blacksquare$$

- note - irrespective of the true value of θ, we know the probability content of $B = \{t : t = T(\theta, x)$ for some $\theta \in A\}$ before we see the data, so is it reasonable to transfer our belief concerning B to A after seeing the data?

- the fiducial probability is not obtained by conditioning or by putting a prior on θ

Example V.1

- suppose $S(x) \in \mathbb{R}$ and let $T(\theta, x) = F_{\theta,S}(S(x))$ be the cdf of S when $\theta \in \mathbb{R}$ is the true value

- when S has a continuous distribution for each θ and $F_{\theta,S}(s)$ is strictly increasing in s, then, when $t \in [0,1]$ we have $P(T(\theta, x) \leq t) = t$ for every θ
(Proof: When $X \sim F$, then
$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u$.)

- suppose in addition that $T(\theta, x)$ is increasing in θ for each x, then for given x

$$\begin{aligned} P_{fiducial,x}(\theta \leq \theta_0) &= P(\{t : T(\theta, x) = t \text{ for some } \theta \leq \theta_0\}) \\ &= T(\theta_0, x) - \inf_{\theta \leq \theta_0} T(\theta, x) \end{aligned}$$

■

- if $A, B \subset \Theta$ and $A \cap B = \phi$, then when $T(\cdot, x)$ is 1-1 for each x

$$
\begin{aligned}
& P_{fiducial,x}(A \cup B) = P(\{t : T(\theta, x) = t \text{ for some } \theta \in A \cup B\}) \\
= \; & P(\{t : T(\theta, x) = t \text{ for some } \theta \in A\} \cup \\
& \quad \{t : T(\theta, x) = t \text{ for some } \theta \in B\}) \\
= \; & P(\{t : T(\theta, x) = t \text{ for some } \theta \in A\}) + \\
& \quad P(\{t : T(\theta, x) = t \text{ for some } \theta \in B\})
\end{aligned}
$$

and so $P_{fiducial}$ is additive but

$$P_{fiducial,x}(\Theta) = P(\{t : T(\theta, x) = t \text{ for some } \theta \in \Theta)\})$$

is not necessarily equal to 1 unless $T(\cdot, x)$ is onto the range of T

- perhaps the easiest context for the fiducial argument is when $\{f_\theta : \theta \in \Theta\}$ is a group family (Fraser's structural inference) recall

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. $P_\theta = P_{\bar{g}(\theta_0)} = P_{\theta_0} \circ g^{-1}$. ■

- we can write a group model algebraically as structural equation

$$x = g(z)$$

where $z \sim P_{\theta_0}$ since $P_{\theta_0}(g(z) \in B) = P_{\theta_0}(g^{-1}B) = P_{\theta_0} \circ g^{-1}(B)$ and we refer to z as the *error distribution*

- note that when the group acts freely on Θ (for any $\theta, \bar{g}_1(\theta) = \bar{g}_2(\theta)$ iff $\bar{g}_1 = \bar{g}_2$) then, after choosing θ_0 we can replace we can replace the parameter space Θ by G since θ identifies g in $\bar{g}(\theta_0)$ and conversely

Definition V.2 For group G acting on $\mathcal{X}$ the set $Orb(x) = \{g(x) : g \in G\}$ is called the orbit of x. ■

Lemma VI.1 The orbits partition $\mathcal{X}$.
Proof: Clearly $\mathcal{X} = \cup_{x \in \mathcal{X}} Orb(x)$ since $x = e(x) \in Orb(x)$. If $x \in Orb(x_1) \cap Orb(x_2)$, then there exists g_1, g_2 s.t. $x = g_1(x_1) = g_2(x_2)$ so $x_1 = g_1^{-1}(g_2(x_2)) = \left(g_1^{-1} \cdot g_2\right)(x_2)$ and so $x_1 \in Orb(x_2)$, which implies $Orb(x_1) \subset Orb(x_2)$ and similarly $Orb(x_2) \subset Orb(x_1)$ so $Orb(x_1) = Orb(x_2)$. ■

- let $orb(x) \in Orb(x)$ be s.t. $orb(g(x)) = orb(x)$ for all $g \in G$ so orb is invariant under G and, assuming G acts freely on $\mathcal{X}$, define $[\cdot] : \mathcal{X} \to G$ by $x = [x](orb(x))$

Lemma VI.2 $[\cdot]$ is equivariant.
Proof: We have

$$g(x) = [g(x)](orb(g(x))) = [g(x)](orb(x))$$

and also

$$g(x) = g([x](orb(x))) = (g \cdot [x])(orb(x))$$

and so by the freeness we must have $[g(x)] = g \cdot [x]$. ■

- so we can write

$$x = [x](orb(x)) = g(z) = g([z](orb(z))) = (g \cdot [z])(orb(z))$$

and since $orb(x) = orb(z)$ this implies

$$[x] = g \cdot [z]$$

and so

$$g = [x] \cdot [z]^{-1} \tag{1}$$

- once we have observed x, then $[x]$ is known, but $[z]^{-1}$ is not known but follows a probability distribution and so (1) induces a probability distribution on g (called the *structural distribution*) and recalled g is just another way to parameterize $\{f_\theta : \theta \in \Theta\}$ so $g \leftrightarrow \theta$

- what probability distribution does $[z]$, and thus $[z]^{-1}$, follow?

- note that by observing x we know $orb(x) = orb(z)$ and so we know something about z and thus $[z]$

Lemma VI.3 *orb* is an ancillary statistic.
Proof: We have for any θ

$$\begin{aligned} P_\theta(orb(x) \in O) &= P_{\bar{g}(\theta_0)}(orb(x) \in O) \\ &= P_{\theta_0}(orb(g(x)) \in O) \\ &= P_{\theta_0}(orb(x) \in O) \end{aligned}$$

and so the distribution of *orb* is independent of θ and fixed by θ_0. ■

- therefore, it is natural to take the structural distribution of $[z]^{-1}$ in (1) to be its conditional distribution under P_{θ_0} given $orb(x)$, namely, for $H \subset G$

$$P_{\theta_0}([z]^{-1} \in H \,|\, orb(x))$$

Example V.2

- $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ iid $N(\mu, 1)$ where $\mu \in \mathbb{R}$ is unknown so $x \sim N_n(\mu 1_n, I)$

- $G = (\mathbb{R}, +)$ is a group acting on $\mathbb{R}^n$ via $g(x) = g1_n + x$ and the action is free

- clearly this group leaves the model invariant and the model is a group model

- we can take the error distribution to be any $N_n(\mu_0 1_n, I)$ but here we choose $z \sim N_n(0, I)$ so we have structural equation

$$x = \mu 1_n + z$$

and we can identify the group here with the parameter

- then

$$\begin{aligned} Orb(x) &= \{\mu 1_n + x : \mu \in \mathbb{R}\} = \{\mu 1_n + \bar{x}1_n + (x - \bar{x}1_n) : \mu \in \mathbb{R}\} \\ &= \{(\mu + \bar{x})1_n + (x - \bar{x}1_n) : \mu \in \mathbb{R}\} \\ &= \{\mu 1_n + (x - \bar{x}1_n) : \mu \in \mathbb{R}\} \end{aligned}$$

and so $x - \bar{x}1_n \in Orb(x)$ and if $y \in Orb(x)$, then for some μ,

$$y = \mu 1_n + x = (\mu + \bar{x})1_n + (x - \bar{x}1_n)$$

so $y - \bar{y}1_n = x - \bar{x}1_n$ so we can take $orb(x) = x - \bar{x}1_n$ and then $[x] = \bar{x}$

- note $[z]^{-1} = -\bar{z}$ so the structural distribution of μ is given by

$$\mu = [x][z]^{-1} = \bar{x} - \bar{z}$$

where $\bar{z}$ has the distribution of the sample average from a sample of n from a $N(0,1)$ given $z - \bar{z}1_n$

- using $1_n \perp z - \bar{z}1_n$

$$\begin{aligned}(2\pi)^{-n/2}\exp\left(-z^t z/2\right) &= (2\pi)^{-n/2}\exp\left(-(\bar{z}1_n + (z-\bar{z}1_n)^t(\cdot)/2\right) \\ &= (2\pi)^{-n/2}\exp\left(-n\bar{z}^2/2\right)\exp(z-\bar{z}1_n)^t(\cdot)/2)\end{aligned}$$

so $\bar{z} \sim N(0, 1/n)$ independent of $z - \bar{z}1_n$

- therefore the structural distribution of μ is $N(\bar{x}, 1/n)$

- note that the structural probability that

$$\mu \in \left[\bar{x} - n^{-1/2} z_{(1+\gamma)/2}, \bar{x} + n^{-1/2} z_{(1+\gamma)/2}\right]$$

is γ and this is also a γ-confidence interval for μ ■

- the correspondence between structural probability and confidence will only happen for parameters that have an induced action under the group, for example, in Ex. VI.2 then μ^2 does not have an induced action

- in such a context is the marginal distribution of μ^2 when $\mu \sim N(\bar{x}, 1/n)$ the correct structural distribution?

- of some interest is that in Ex. VI.2 if we change the family of distributions to say $x = \mu 1_n + z$ where z is a sample of n from any absolutely continuous distribution on $\mathbb{R}$, then the same analysis goes through except the conditional distribution of $\bar{z}$ given $z - \bar{z}1_n$ will depend on $z - \bar{z}1_n$

- basically with any group G acting freely on $\mathcal{X}$ you can take any distribution P on $\mathcal{X}$ for the error $z \sim P$ and then $x = g(z) \sim P_g = P \circ g^{-1}$

- there are many examples of group models where the structural argument applies such as linear regression models

- one interesting aspect of structural inference is that it is also equivalent to an improper Bayesian analysis (to be discussed)

- one issue is that sometimes a model can be represented in different ways and then structural inferences of the same parameter can be different, this is known as the *marginalization paradox*

Example V.3

- suppose $x \sim N_k(\mu, \Sigma)$ and write (spectral decomposition) $\Sigma = Q\Lambda Q^t$ where Q is $k \times k$ orthogonal, Λ is $k \times k$ diagonal with positive elements on the diagonal

- then, using Gram-Schmidt on the columns in the order first to last, we can write $\Sigma^{1/2} = Q\Lambda^{1/2}Q^t = U_1 T_\nabla$ where U_1 is $k \times k$ orthogonal, and T_∇ is $k \times k$ upper triangular with positive elements on the diagonal and so $\Sigma = \left(\Sigma^{1/2}\right)^t \Sigma^{1/2} = T_\nabla^t T_\nabla = T_\triangle T_\triangle^t$ where $T_\triangle = T_\nabla^t$ is lower triangular with positive diagonal (Cholesky factorization of Σ)

- but we can also do the Grahm-Schmidt process on in the reverse order (last column to first) to obtain $\Sigma^{1/2} = U_2 T_\triangle$ where U_2 is $k \times k$ orthogonal, and $T_\triangle$ is $k \times k$ lower triangular with positive elements on the diagonal and so $\Sigma = \left(\Sigma^{1/2}\right)^t \Sigma^{1/2} = T_\triangle^t T_\triangle = T_\nabla T_\nabla^t$ where T_∇ is upper triangular with positive diagonal

- the model $\{N_k(\mu, \Sigma) : \mu \in \mathbb{R}^k, \Sigma \in \mathbb{R}^{k\times k}$ p.d.$\}$ is a group model via the action

$$x = (\mu, T_\triangle)(z) = \mu + T_\triangle z$$

where $z \sim N_k(0, I)$ and $G_\triangle = \{(\mu, T\triangle) : \mu \in \mathbb{R}^k, T_\triangle \in \mathbb{R}^{k\times k}$ lower triangular with positive elements on the diagonal$\}$ with product

$$(\mu_1, T_{\triangle 1}) \cdot (\mu_2, T_{\triangle 21}) = (\mu + T_{\triangle 1}\mu_2, T_{\triangle 1} T_{\triangle 21})$$

and note that for a sample $x = (x_1, \ldots, x_n)$ the group acts coordinate-wise as $g(x) = (g(x_1), \ldots, g(x_n))$
- but it can also be written as

$$x = (\mu, T_\nabla)(z) = \mu + T_\nabla z$$

where $z \sim N_k(0, I)$ and $G_\nabla = \{(\mu, T_\nabla) : \mu \in \mathbb{R}^k, T_\nabla \in \mathbb{R}^{k\times k}$ upper triangular with positive elements on the diagonal$\}$ with product as above

- using $G_{\triangle}$, then for a sample $[\bar{x}, S_{x\triangle}]$, where $S_x =$ (n-1) times the sample covariance matrix, is equivariant and $orb(x) = [\bar{x}, S_{x\triangle}]^{-1}x$ which induces structural distribution

$$(\mu, T_{\triangle}) = [\bar{x}, S_{x\triangle}][\bar{z}, S_{z\triangle}]^{-1}$$

based on the distribution of $(\bar{z}, S_{z\triangle})$ (this is independent of $orb(x)$) and this in turn induces a distribution on $\Sigma = T_{\triangle}T_{\triangle}^t$

- we can do the same thing but using G_{∇} and problem is (marginalization paradox) that the induced distribution on $\Sigma = T_{\nabla}T_{\nabla}^t$ is not the same as that obtained using $G_{\triangle}$

- so which group is correct (if any)?

- Fraser would argue that the group structure is an explicit part of the model and so one should not be surprised by different answers, but the difficulty is how to see what determines this in an application

- this is also a problem for the improper prior approach as the different groups lead to different improper priors and so different inferences ■

- there has been, and there still is, considerable research activity around fiducial such as generalized fiducial (see the work of Jan Hannig), confidence distributions (see the work of Minge Xie), belief fuctions (see the work of Art Dempster) and inferential models /imprecise probability (see the work of Ryan Martin)

- while there is lots going on in this area, it doesn't seem to me that the basic issues associated with fiducial have been resolved, but perhaps in the future they will be