# 9. Invariance

## (a) Principle of invariance

- data $x = (x_1, \ldots, x_n) \overset{iid}{\sim} \{f_\theta : \theta \in \Theta\}$
- inferences about $\psi = \Psi(\theta)$
- suppose $Y : \mathcal{X} \to \mathcal{Y}$ (Euclidean space) 1-1, smooth
  so $y = (y_1, \ldots, y_n) = (Y(x_1), \ldots, Y(x_n)) \sim \{g_\theta : \theta \in \Theta\}$
  where $g_\theta(y) = f_\theta(Y^{-1}(y)) J_{Y^{-1}}(y)$
  $J_Y(y) = \left|\det\left(\frac{\partial Y_i}{\partial x_j}(Y^{-1}(y))\right)\right|$
- also let $\Lambda : \Psi \to \Lambda$ be 1-1, smooth s.t.
  $\lambda = \Lambda(\Psi(\theta))$
- principle of invariance
  Inferences about the object of interest should be invariant under smooth 1-1 transformations.

- eg - if we estimate $\psi$ by MLE then we should obtain the same estimate if we base the estimate on $y$ instead and give the same accuracy

  - if we choose to estimate $\lambda$ instead of $\psi$ then, based on data $x$, we should have $\lambda(x) = \Lambda(\psi(x))$ and have the same accuracy

  - if there is evidence in favor of (against) $\psi_0$ there must be evidence in favor of (against) $\lambda_0 = \Lambda(\psi_0)$

and the evidence must be of the same strength whether using $x$ or $y$.

- in general decision theory (and much of inference) does not satisfy the invariance principle but there are approaches to inference that do (later)

- a restricted notion of invariance is used.

note - consider a set $\Theta$ and the set $\mathcal{F}$ of all functions $f : \Theta \xrightarrow[\text{onto}]{1-1} \Theta$ and s.t. $f, f^{-1}$ satisfy the same smoothness conditions
- then if $f, g \in \mathcal{F}$ then $f \circ g \in \mathcal{F}$ (closure)
(i) $(f \circ g) \circ h = f \circ (g \circ h)$ $\forall f, g, h \in \mathcal{F}$ (associative)
(ii) $e \in \mathcal{F}$ where $e(\theta) = \theta$ $\forall \theta \in \Theta$ (identity)
(iii) $f \circ f^{-1} = f^{-1} \circ f = e$ $\forall f \in \mathcal{F}$ (inverse)

- a set $G$ with binary operation $\circ : G \times G \to G$ is called a group when $\circ$ satisfies (i), (ii), (iii)

eg the set of all smooth reparameterizations on a set $\Theta$ is a group

eg location group $(\mathbb{R}, +)$

eg location-scale group $(a, c) \in \mathbb{R} \times (0, \infty)$ with $(a_1, c_1) \circ (a_2, c_2) = (a_1 + c_1 a_2, c_1 c_2)$

$e = (0, 1)$, $(a, c)^{-1} = (-a/c, 1/c)$

**eg** matrix groups

$G = \{C \in \mathbb{R}^{k \times k} : C \text{ invertible}\}$

$G = \{C \in \mathbb{R}^{k \times k} : C \text{ upper } \Delta \text{ with positive diagonal}\}$

$G = \{C \in \mathbb{R}^{k \times k} : C \text{ orthogonal}\}$

+ location-scale version $(\underset{\sim}{a}, C) \in \mathbb{R}^k \times \mathbb{R}^{k \times k}$ with
$(\underset{\sim}{a}_1, C_1)(\underset{\sim}{a}_2, C_2) = (\underset{\sim}{a}_1 + C_1 \underset{\sim}{a}_2, C_1 C_2)$

- transformation groups: $G$ a group and for each $g \in G$ $\exists$ $T_g : \mathcal{X} \xrightarrow{1-1} \mathcal{X}$ satisfying $T_{g_1 g_2} = T_{g_1} \circ T_{g_2}$ and $T_e$ = identity on $\mathcal{X}$ and we say $G$ acts on $\mathcal{X}$

- if $G$ acts as above then $(\{T_g : g \in G\}, \circ)$ is a group with $T_g^{-1} = T_{g^{-1}}$.

**eg** location group acting on $\mathbb{R}^n$

$T_a x = a\underset{\sim}{1} + x \quad \text{for } a \in \mathbb{R}$

**eg** location-scale group acting on $\mathbb{R}^n$

$T_{(a,c)} x = a\underset{\sim}{1} + cx \qquad \underset{\sim}{1} = (1, \ldots, 1)'$

**eg** matrix group acting on $\mathbb{R}^{k \times n}$

$T_{(\underset{\sim}{a}, C)} x = \underset{\sim}{a}\underset{\sim}{1}' + Cx$

- <u>note</u> if $G$ acts on $\mathcal{X}$ then it also acts on $\mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}$ via $T_g(x_1, \ldots, x_n) = (T_g x_1, \ldots, T_g x_n)$

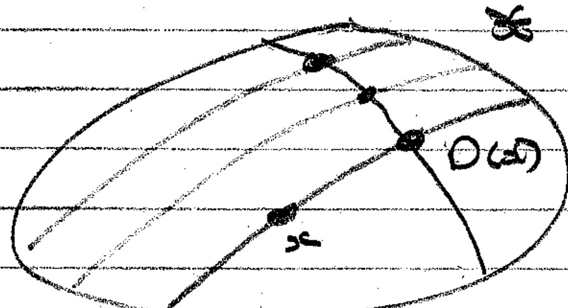
- for $G$ acting on $\mathcal{X}$ the set $\{T_g x : g \in G\}$ is called the <u>orbit</u> of $x$.

- a function $D : \mathcal{X} \to \mathcal{X}$ satisfying $D(T_g x) = D(x)$ $\forall g \in G, \forall x \in \mathcal{X}$ is an invariant and is a <u>maximal invariant</u> when $D(x_1) = D(x_2)$ iff $x_1$ and $x_2$ are in the same orbit

- note the orbits partition $\mathcal{X}$ ($x_1$ and $x_2$ are in the same orbit iff $\exists\, g$ s.t. $x_1 = T_g x_2$) and a maximal invariant indexes orbits

- we can always "define" a maximal invariant by picking (axiom of choice) from each orbit and defining $D(x)$ to be that element

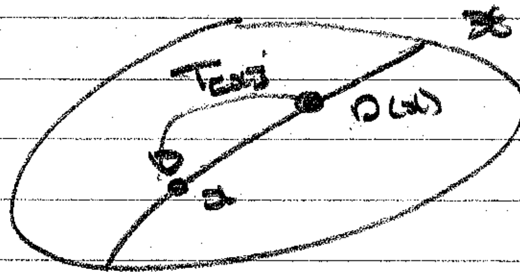
but we want to do this smoothly




- $G$ acts <u>freely</u> on $\mathcal{X}$ when $T_{g_1} x = T_{g_2} x$ iff $g_1 = g_2$ (sometimes this requires the removal of a set of volume measure 0 from $\mathcal{X}$)

transformation variable 

- when $G$ acts freely and $D$ is a maximal invariant then define $[\cdot] : \mathcal{X} \to G$ by $x = T_{[x]} D(x)$




and the freeness implies $[\cdot]$ is well-defined.

- when $G$ acts freely with max. invariant $D$ then $[\cdot]$ satisfies $[T_g x] = g[x]$

Proof: $T_g x = T_g T_{[x]} D(x) = T_{[T_g x]} D(T_g x)$

$= T_{[T_g x]} D(x)$ and so $T_g T_{[x]} = T_{g[x]} = T_{[T_g x]}$

which proves the result.

- if we have $[\cdot] : \mathcal{X} \to G$ satisfying $[T_g x] = g[x]$ $\forall g \in G$, $x \in \mathcal{X}$ then $D(x) = T_{[x]}^{-1} x = T_{[x]^{-1}} x$ is a maximal invariant

Proof: Suppose $x' = T_g x$ then $D(x') = D(T_g x)$

$= T_{[T_g x]^{-1}} T_g x = T_{[x]^{-1} g^{-1}} T_g x = T_{[x]^{-1}} T_{g^{-1}} T_g x$

$= T_{[x]^{-1}} x = D(x)$. If $x_1$ and $x_2$ are in

different orbits iff $T_g x_1 \neq x_2$ $\forall g$ and if

$D(x_1) = D(x_2)$ then $x_1 = T_{[x_1]} D(x_1) = T_{[x_1]} D(x_2)$

$= T_{[x_1]} T^{-1}_{[x_2]} T_{[x_2]} D(x_2) = T_{[x_1][x_2]^{-1}} x_2$ (*).

- notation - $X : \Omega \to \mathfrak{X}$, $P$ a prob measure on $\Omega$ and $P_X(A) = P(X^{-1}B)$

- $T : \mathfrak{X} \xrightarrow[\text{smooth}]{1-1} \mathcal{Y}$ and $Y = T(X)$

- $P_Y(A) = P(Y^{-1}A) = P(X^{-1}T^{-1}A)$

$= P_X(T^{-1}A) = P_X \circ T^{-1}(A)$ notation

- also if $X$ has density $f_X$ wrt $\mu$ and $T$ is sufficiently smooth then $Y$ has density

$$f_Y(y) = f_X(T^{-1}(y)) J_T(T^{-1}(y))$$

$$J_T(x) = |\det dT(x)|^{-1} = \left|\det\left(\frac{\partial T_i(x)}{\partial x_j}\right)\right|^{-1}$$

- also useful $J_{T^{-1}}(y) = |\det dT^{-1}(y)|$

inverse fn thm

$= |\det(dT(T^{-1}(y)))^{-1}|^{-1}$

$= |\det(dT(T^{-1}(y)))| = J_T^{-1}(T^{-1}(y))$

$\therefore J_T(T^{-1}(y)) = J^{-1}_{T^{-1}}(y)$ and most convenient approach can be chosen for computations

$$J_{T_1 \circ T_2}(x) = |\det(dT_1 \circ T_2)(x)|^{-1}$$

chain rule

$$= |\det(dT_1(T_2(x))\, dT_2(x))|^{-1}$$

$$= J_{T_1}(T_2(x))\, J_{T_2}(x)$$

- also if $X \sim P$ and $h : \mathcal{Y} \to \mathbb{R}$ then

$$E_{P_Y}(h) = \int_{\mathcal{Y}} h(y)\, P_Y(dy)$$

$$= \int_{\mathcal{Y}} h(y)\, P \circ T^{-1}(dy)$$

$$= \int_{\mathcal{X}} h(T(x))\, P(dx)$$

indicators simple fns etc.

(b) Symmetries of a Model

- $G$ acting on $\mathcal{X}$ is a symmetry group of model $\{P_\theta : \theta \in \Theta\}$ if $P_\theta \circ T_g^{-1} \in \{P_\theta : \theta \in \Theta\}$ $\forall \theta \in \Theta, g \in G$ (leaves model invariant)

Lemma ① If $G$ is a symmetry group of the model then $G$ acts on $\Theta$ via $\bar{T}_g \theta = \theta'$ when $P_{\theta'} = P_\theta \circ T_g^{-1}$.

Proof: Since $\theta$ indexes $\bar{T}_g \theta$ is well-defined

If $\bar{T}_g \theta_1 = \bar{T}_g \theta_2$, then $P_{\theta_1} \circ T_g^{-1} = P_{\theta_2} \circ T_g^{-1}$ so $P_{\theta_1}(T_g^{-1} B) = P_{\theta_2}(T_g^{-1} B)$ $\forall B$ so $P_{\theta_1} = P_{\theta_2}$ since $\theta$ indexes and $\bar{T}_g : \Theta \to \Theta$ is 1-1.

Also $P_{\bar{\bar{T}}_{g_1 g_2}\theta} = P_\theta \circ T_{g_1 g_2}^{-1} = P_\theta \circ (T_{g_1} \circ T_{g_2})^{-1}$
$= P_\theta \circ T_{g_2}^{-1} \circ T_{g_1}^{-1} = P_{\bar{\bar{T}}_{g_2}\theta} \circ T_{g_1}^{-1} = P_{\bar{\bar{T}}_{g_1}\bar{\bar{T}}_{g_2}\theta} \quad \forall \theta$
and so $\bar{\bar{T}}_{g_1 g_2} = \bar{\bar{T}}_{g_1} \bar{\bar{T}}_{g_2}$. Finally $P_{\bar{\bar{T}}_e \theta}$
$= P_\theta \circ T_e^{-1} = P_\theta$ and so $\bar{\bar{T}}_e \theta = \theta \ \forall \theta$.

__Corollary__ ② $(\{\bar{\bar{T}}_g : g \in G\}, \circ)$ is a group.

__eg__ $\{N(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

- if $x \sim N(\mu, \sigma^2)$ then $y = T_{(a,c)}(x)$
  $\sim N(a + c\mu, c^2\sigma^2)$

- so location-scale group is a symmetry group of this model and $\bar{\bar{T}}_{(a,c)}(\mu, \sigma^2) = (a + c\mu, c^2\sigma^2)$

__eg__ $\{N_k(\underline{\mu}, \Sigma) : \underline{\mu} \in \mathbb{R}^k, \Sigma \text{ pd.}\}$

- $\underline{a} \in \mathbb{R}^k$, $C \in \mathbb{R}^{k \times k}$ upper $\Delta$ with pos. diagonal.

- if $\underline{x} \sim N_k(\underline{\mu}, \Sigma)$ then $\underline{y} = T_{(\underline{a},C)}\underline{x}$
  $\sim N_k(\underline{a} + C\underline{\mu}, C\Sigma C')$ and so
  $\bar{\bar{T}}_{(\underline{a},C)}(\underline{\mu}, \Sigma) = (\underline{a} + C\underline{\mu}, C\Sigma C')$

eg <u>structural or group models</u> (Fraser)

- suppose $G$ acts <u>transitively</u> on $\Theta$, namely, for any $\theta_1, \theta_2 \in \Theta$ $\exists$ $g$ s.t. $T_g\theta_1 = \theta_2$

- when $G$ acts transitively the orbit of $\theta$ is $\{T_g\theta : g \in G\} = \Theta$

- also suppose $G$ acts freely on $\Theta$

- pick $\theta_0 \in \Theta$ and suppose $z \sim P_{\theta_0}$

- then $x = T_g z \sim P_{\bar{T}_g\theta_0}$ and for each $\theta \in \Theta$ $\exists!$ $g \in G$ s.t. $\bar{T}_g\theta_0 = \theta$

- so $x = T_g z$ with $z \sim P_{\theta_0}$ is an algebraic representation of the model and instead of making inference about the true $\theta \in \Theta$ we can "equivalently" make inference about the true $g \in G$.

- alternatively we can start with $z \sim P_e$ and group $G$ acting on $\mathcal{X}$ and put $x = T_g z \sim P_{\bar{T}_g e} = P_g$ then if $P_g \neq P_{g'}$ $\forall g, g' \in G$ we have a <u>group model</u>

<u>eg</u> $\{N_k(\underset{\sim}{\mu}, \Sigma) : \underset{\sim}{\mu} \in \mathbb{R}^k, \Sigma \text{ p.d.}\}$

- consider $G = \{(\underset{\sim}{a}, C) : \underset{\sim}{a} \in \mathbb{R}^k, C \text{ upper}\,\Delta \text{ pos. diag}\}$

- $e = (0, I)$ so put $P_e = N_k(\underset{\sim}{0}, I)$ and if $\underset{\sim}{z} \sim N_k(\underset{\sim}{0}, I)$ then

$$\underset{\sim}{y} = T_{(\underset{\sim}{a}, C)} \underset{\sim}{z} \sim N_k(\underset{\sim}{a}, CC')$$

- let $\Sigma \underset{\text{decomp}}{\overset{\text{spectral}}{=}} Q \Lambda Q' = Q \Lambda^{1/2} Q' Q \Lambda^{1/2} Q'$

$$= \Sigma^{1/2} \Sigma^{1/2} \quad \text{(symmetric square root)}$$

$\Sigma^{1/2} = UC', \; U \in \mathbb{R}^{k \times k}, \; C' \in \mathbb{R}^{k \times k}$ lower $\Delta$ with pos. diagonal

via Gram-Schmidt on the columns of $\Sigma^{1/2}$ last column to first

$$\Sigma = (\Sigma^{1/2})' \Sigma^{1/2} = CC' \text{ and } C \text{ is unique}$$
(Cholesky factorization)

and denote it $\Sigma_0$

- so $\underset{\sim}{y} = T_{(\underset{\sim}{\mu}, \Sigma_0)}(\underset{\sim}{z}) \sim N_k(\underset{\sim}{\mu}, \Sigma)$

- the action on $\Theta = \{(\underset{\sim}{\mu}, \Sigma) : \underset{\sim}{\mu} \in \mathbb{R}^k, \Sigma \text{ p.d.}\}$ is transitive since

$(\underset{\sim}{\mu}_2, \Sigma_2) = \overline{T}_{(a,C)}(\underset{\sim}{\mu}_1, \Sigma_1) = (\underset{\sim}{a} + C\underset{\sim}{\mu}_1, C\Sigma_1 C')$

so put $C = \Sigma_{20} \Sigma_{10}^{-1}$, $\underset{\sim}{a} = \underset{\sim}{\mu}_2 - C\underset{\sim}{\mu}_1$

and it is free since

$$\bar{T}_{(\underset{\sim}{a}_1, C_1)}(\underset{\sim}{\mu}, \Sigma) = (\underset{\sim}{a}_1 + C_1 \underset{\sim}{\mu}, C_1 \Sigma C_1')$$

$$= (\underset{\sim}{a}_2 + C_2 \underset{\sim}{\mu}, C_2 \Sigma C_2') = \bar{T}_{(\underset{\sim}{a}_2, C_2)}(\underset{\sim}{\mu}, \Sigma)$$

iff $C_1 \Sigma_\Delta \Sigma_\Delta' C_1' = C_2 \Sigma_\Delta \Sigma_\Delta' C_2'$

$\underset{\sim}{a}_1 + C_1 \underset{\sim}{\mu} = \underset{\sim}{a}_2 + C_2 \underset{\sim}{\mu}$

iff $C_1 \Sigma_\Delta = C_2 \Sigma_\Delta$, $\underset{\sim}{a}_1 + C_1 \underset{\sim}{\mu} = \underset{\sim}{a}_2 + C_2 \underset{\sim}{\mu}_2$

iff $C_1 = C_2$, $\underset{\sim}{a}_1 = \underset{\sim}{a}_2$.

- so can write this as a group model
  $\underset{\sim}{x} = \bar{T}_{(\underset{\sim}{\mu}, \Sigma_\Delta)} \underset{\sim}{z}$ with $\underset{\sim}{z} \sim N_n(\underset{\sim}{0}, I)$

- <u>note</u> - this representation is not unique
  as we don't have to take $\underset{\sim}{z} \sim N_n(\underset{\sim}{0}, I)$

  - but also $\underset{\sim}{x} = \bar{T}_{(\underset{\sim}{\mu}, \Sigma_\Delta)} \underset{\sim}{z} = \underset{\sim}{\mu} + \Sigma_\Delta \underset{\sim}{z}$ with $\Sigma_\Delta$ lower $\Delta$ with positive diagonal is a valid representation.

eg <u>multivariate Cauchy</u>

- $\underset{\sim}{x} = \bar{T}_{(\underset{\sim}{\mu}, \Sigma_\Delta)} \underset{\sim}{z}$ with $\underset{\sim}{z} \sim \dfrac{\Gamma\left(\frac{k+1}{2}\right)}{\pi^{\frac{k+1}{2}}} \left(\dfrac{1}{1 + \underset{\sim}{z}'\underset{\sim}{z}}\right)^{-\frac{k+1}{2}}$

- note - $\underset{\sim}{\mu}$ is the mode and $\Sigma$ prescribes shape of ellipsoidal contours but is not the variance matrix.

**Lemma ③** If $G$ is a symmetry group of $\{P_\theta : \theta \in \Theta\}$ and $D$ is a max. invariant then the distribution of $D$ depends on $\theta$ only through a maximal invariant $\bar{D}$ on $\Theta$.

Proof: $P_\theta(D(x) \in A) = P_\theta(D(T_g x) \in A)$
$= P_{\bar{T}_g \theta}(D(x) \in A)$ and so the distribution of $D$ is constant for $\theta' \in \{\bar{T}_g \theta : g \in G\}$.

**Corollary ④** If $G$ acts transitively on $\Theta$ then $D$ is ancillary.

Proof: $\Theta$ is the only orbit

- note - if a fn $h$ defined on $\mathcal{X}$ is invariant under $G$, i.e. $h(T_g x) = h(x)\ \forall x, g$ then $h$ is constant on orbits of $G$ on $\mathcal{X}$ and so is a function of a maximal invariant $D(x)$, $h(x) = k(D(x))$ for some fn $k$ and conversely

- $h : \mathcal{X} \xrightarrow{\text{onto}} \mathcal{H}$ is equivariant if whenever $h(x_1) = h(x_2)$ then $h(T_g x_1) = h(T_g x_2)$ and then we can define an action on $\mathcal{H}$ via $T_{\mathcal{H}g}\, h(x) = h(T_g x)$

<u>eg group models</u>

- $G$ acts transitively and freely on $\Theta$

- also assume $G$ acts freely on $\mathcal{X}$ (perhaps after removing a set of support measure 0)

- let $[\cdot]$ be a transformation variable with associated maximal invariant $D$

- then the stat. model can be represented as $x = T_g z$ where $z \sim f_\Theta$ for some $f_\Theta$ in the model

- also $x = T_{[x]} D(x) = T_g z = T_g T_{[z]} D(z)$ and since $D(x) = D(z)$ we must have $[x] = g[z]$ and we want to make inference about $g \in G$

- since $D$ is ancillary the relevant distribution of $[z]$ is the conditional $[z] \mid D(z) = D(x)$

- then $g = [x][z]^{-1}$ and since $[x]$ is fixed and known this places a probability distribution on $g$ known as the structural distribution

- the quantity $[z] = g^{-1}[x]$ is known as a <u>pivotal</u> (a function of data and parameter whose distribution is fixed)

- more generally pivotals give rise to distributions on parameters known as fiducial distributions

- fiducial distributions, unlike posteriors, do not arise via an application of conditional probability and are somewhat outside the realm of probability.

- fiducial is appealing to some because it provides probability distributions on parameters without the need to specify a prior.

- see section (d) for examples.

(c) <u>Distribution theory for group models</u>

- <u>note</u> - all densities here are with respect to volume measure (length, area, etc.) on respective spaces which have appropriate structure (Riemann manifolds)

  - when space is discrete, volume measure = counting measure

- Suppose we have a group model where

$$x = T_g z$$

where $z \sim f_e$ (some fixed density)

- then $x$ has density

$$f_g(x) = f_e(T_g^{-1}x) J_{T_g}(T_g^{-1}x)$$

- let $U : \mathcal{X} \xrightarrow{1-1} G \times D$ be given by $U(x) = ([x], D(x))$ then for $(h,d) \in G \times D$

$$f_{U,g}(h,d) = f_g(U^{-1}(h,d)) J_U(U^{-1}(h,d))$$

$$= f_g(T_h d) J_U(T_h d)$$

$$= f_e(T_g^{-1}T_h d) J_{T_g}(T_g^{-1}T_h d) J_U(T_h d)$$

$\propto f_g(h \mid d)$ the conditional of $h \mid \underline{d}$

eg location models

- $\{ f_0(\underset{\sim}{x} - \mu \underset{\sim}{1}) : \mu \in \mathbb{R} \}$ same density $f_0$

- $G = (\mathbb{R}, +)$ and for $\underset{\sim}{x} \in \mathbb{R}^n$ then $T_g \underset{\sim}{x} = g\underset{\sim}{1} + \underset{\sim}{x}$ is the action

- $[\underset{\sim}{x}] = \bar{x}$ satisfies $[T_g \underset{\sim}{x}] = [g\underset{\sim}{1} + \underset{\sim}{x}] = g + \bar{x}$ and so is a transformation variable with max. invariant $D(\underset{\sim}{x}) = T^{-1}_{[\underset{\sim}{x}]} \underset{\sim}{x} = \underset{\sim}{x} - \bar{x}\underset{\sim}{1} = \underset{\sim}{d}$

- note that $\underset{\sim}{d}'\underset{\sim}{1} = 0$ so $\underset{\sim}{d} \in \mathcal{L}^{\perp}\{\underset{\sim}{1}\}$

- also $dT_g(\underset{\sim}{x}) = \left( \dfrac{\partial T_{g_i}(\underset{\sim}{x})}{\partial x_j} \right) = I$

so $J_{T_g}(\underset{\sim}{x}) = 1$

- also for $U(\underset{\sim}{x}) = (\bar{x}, \underset{\sim}{d})$ we have from (taking free coords of $\underset{\sim}{d}$ to be $d_1, \ldots, d_{n-1}$ as $d_1 + \cdots + d_n = 0$) $\underset{\sim}{x} = \bar{x}\underset{\sim}{1} + \underset{\sim}{d}$

$$dU(\bar{x}, d_1, \ldots, d_{n-1})$$

$$= \begin{pmatrix} \frac{\partial x_1}{\partial \bar{x}} & \frac{\partial x_1}{\partial d_1} & \cdots & \frac{\partial x_1}{\partial d_{n-1}} \\ \vdots & & & \\ \frac{\partial x_n}{\partial \bar{x}} & \frac{\partial x_n}{\partial d_1} & \cdots & \frac{\partial x_n}{\partial d_{n-1}} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ & & & & 1 \\ 1 & -1 & -1 & \cdots & -1 \end{pmatrix}$$

using $x_n = \bar{x} + d_n = \bar{x} - d_1 - \cdots - d_{n-1}$

and so $J_u(\bar{x}, d_1, \ldots, d_{n-1})$ is constant

- so by the formula on p. 137

$$f_{u,g}(\bar{x} \mid \underset{\sim}{d}) \propto f_0\left((\bar{x}-g)\underset{\sim}{1} + \underset{\sim}{d}\right)$$

- for example, when $f_0 = N_n(\underset{\sim}{0}, I)$, $h = \bar{x}$, $g = \mu$ then (so we have a location normal model)

$$f_{u,\mu}(\bar{x} \mid \underset{\sim}{d}) \propto \exp\left\{-\tfrac{1}{2}\left((\bar{x}-\mu)\underset{\sim}{1} + \underset{\sim}{d}\right)'\left(\;\right)\right\}$$

$$= \exp\left\{-\frac{n}{2}(\bar{x}-\mu)^2\right\}$$

and $\bar{x} \sim N(\mu, \frac{1}{n})$ stat. ind of $\underset{\sim}{d}$.

- for other choices of $f_0$ the stat. ind of $\bar{x}$ and $\underset{\sim}{d}$ will not hold, generally.

eg <u>location-scale</u> $\{ \sigma^{-n} f_{(0,1)}(\frac{x-\mu 1}{\sigma}) : \mu \in \mathbb{R}, \sigma > 0 \}$

- $G = (\mathbb{R} \times (0,\infty), \cdot)$ where $(a_1, c_1)(a_2, c_2) = (a_1 + c_1 a_2, c_1 c_2)$

- $d T_g x = d T_{(a,c)} x = \left( \frac{\partial a + c x_i}{\partial x_j} \right) = c I$

  $\therefore J_{T_g}(x) = c^{-n}$

- $[x] = (\bar{x}, \| x - \bar{x} 1 \|)$ satisfies

  $[T_g x] = (a + c\bar{x}, c \| x - \bar{x} 1 \|) = (a,c)(\bar{x}, \| x - \bar{x} 1 \|)$

  and so $[\cdot]$ is a transformation variable

- then $d = D(x) = T^{-1}_{[x]} x = T_{[x]^{-1}} x$

  $= T_{(-\bar{x}/\|x - \bar{x}1\|, 1/\|x - \bar{x} 1\|)} x$

  $= \frac{x - \bar{x} 1}{\| x - \bar{x} 1 \|}$

  and note that $d'1 = 0$ and $d'd = 1$ so $d$ has $n-2$ free coordinates.

- for $U(x) = (\bar{x}, \| x - \bar{x} 1 \|, d)$

- to simplify notation let $s = \| x - \bar{x} 1 \|$ so

  $x = \bar{x} 1 + s d$ and treating $d_1, \ldots, d_{n-2}$ as free.

$$\det\begin{pmatrix} \frac{\partial x_1}{\partial \bar{x}} & \frac{\partial x_1}{\partial s} & \frac{\partial x_1}{\partial d_1} & \cdots & \frac{\partial x_1}{\partial d_{n-2}} \\ \vdots & & & & \vdots \\ \frac{\partial x_n}{\partial \bar{x}} & \frac{\partial x_n}{\partial s} & \frac{\partial x_n}{\partial d_1} & \cdots & \frac{\partial x_n}{\partial d_{n-2}} \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & d_1 & s & 0 & \cdots & 0 \\ 1 & d_2 & 0 & s & \cdots & 0 \\ \vdots & & & & & \vdots \\ 1 & d_{n-2} & 0 & \cdots & \cdots & s \\ 1 & d_{n-1} & s\,\partial d_{n-1}/\partial d_1 & \cdots & & s\,\partial d_{n-1}/\partial d_{n-2} \\ 1 & d_n & s\,\partial d_n/\partial d_1 & \cdots & & s\,\partial d_n/\partial d_{n-2} \end{pmatrix}$$

$= s^{n-2}\, k(\underset{\sim}{d})$ since $d_{n-1}+d_n = -(d_1+\cdots+d_{n-2})$
and $d_{n-1}^2+d_n^2 = 1-(d_1^2+\cdots+d_{n-2}^2)$
implies $\partial d_{n-1}/\partial d_i + \partial d_n/\partial d_i = -1$
$2d_{n-1}\,\partial d_{n-1}/\partial d_i + 2d_n\,\partial d_n/\partial d_i = -2d_i$
for $i=1,\ldots,n-2$ and this implies

$$\begin{pmatrix} 1 & 1 \\ d_{n-1} & d_n \end{pmatrix}\begin{pmatrix} \partial d_{n-1}/\partial d_i \\ \partial d_n/\partial d_i \end{pmatrix} = \begin{pmatrix} -1 \\ -d_i \end{pmatrix}$$

so $\partial d_{n-1}/\partial d_i$ and $\partial d_n/\partial d_i$ are functions of $d_1,\ldots,d_{n-2}$

- therefore $f_{U,(\mu,\sigma)}(h\,|\,\underset{\sim}{d}) = f_{(\mu,\sigma)}((\bar{x},s)\,|\,\underset{\sim}{d})$

$$\propto f_{(0,1)}\left(\frac{\bar{x}\underset{\sim}{1}+s\underset{\sim}{d}-\mu\underset{\sim}{1}}{\sigma}\right)\sigma^{-n}\, s^{n-2}$$

- so when $f_0$ is $N_n(\underset{\sim}{0}, I)$ we get the joint density for $(\bar{x}, s)$ is $\propto$

$$\sigma^{-n} \exp\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2 - \frac{s^2}{2\sigma^2}\} s^{n-2}$$

which implies $\bar{x} \mid \underset{\sim}{d} \sim N(\mu, \sigma^2/n)$ stat. ind of $s \sim \sigma^2$ chi$(n-1)$ (Exercise) which are stat. ind of $\underset{\sim}{d}$ (Fact: $\underset{\sim}{d} \sim$ uniformly on the unit sphere in $\mathcal{L}^\perp\{\underset{\sim}{1}\}$)

- for general $f_{(0,1)}$ we won't have the mut. stat. ind. of $\bar{x}, s, \underset{\sim}{d}$

## (d) Invariance and decisions

(i) suppose $G$ is a symmetry group of the model $\{f_\theta : \theta \in \Theta\}$

- we have object of interest $\psi = \Psi(\theta)$ and loss $L(\theta, \psi)$

- we say $\Psi$ is equivariant under $G$ when $\Psi(\bar{T}_g \theta_1) = \Psi(\bar{T}_g \theta_2)$ iff $\Psi(\theta_1) = \Psi(\theta_2)$ for any $g \in G$, $\theta_1, \theta_2 \in \Theta$

(ii) when $\Psi$ is equivariant we can define an action on (space) $\Psi$ by $T_g^* \psi = \Psi(\bar{T}_g \theta)$

Proof: Since $\Psi : \Theta \to \Psi$ is onto and $\Psi(\bar{T}_g \theta_1) = \Psi(\bar{T}_g \theta_2)$ iff $\Psi(\theta_1) = \Psi(\theta_2)$ this action is well-defined. Also if $\psi = \Psi(\theta)$, then $T^*_{gh} \psi = \Psi(\bar{T}_{gh}\theta)$ $= \Psi(\bar{T}_g \bar{T}_h \theta) = T_g^* \Psi(\bar{T}_h \theta) = T_g^* T_h^* \Psi(\theta)$ $= T_g^* T_h^* \psi$. Also $T_e^* \psi = \Psi(\bar{T}_e \theta)$ $= \Psi(\theta) = \psi$.

(iii) in addition suppose the loss function satisfies $L(\bar{T}_g \theta, T_g^* \psi) = L(\theta, \psi)$ $\forall g \in G, \theta \in \Theta, \psi \in \Psi$

- under (i), (ii) and (iii) we say the <u>decision problem is invariant</u> under $G$

- suppose now $\delta \in \mathcal{D}$ is a decision fn
- then $G$ acts on $\mathcal{D}$ as follows $\tilde{T}_g \delta(x, A) = \delta(T_g^{-1}x, (T_g^*)^{-1}A)$

Proof: Clearly $\tilde{T}_g \delta \in \mathcal{D}$, $\tilde{T}_e \delta(x, A) = \delta(x, A)$
and $\tilde{T}_{gh} \delta(x, A) = \delta(T_{gh}^{-1}x, (T_{gh}^*)^{-1}A)$
$= \delta(T_h^{-1}T_g^{-1}x, T_h^{*-1}T_g^{*-1}A)$
$= \tilde{T}_h \delta(T_g^{-1}x, T_g^{*-1}A) = \tilde{T}_g \tilde{T}_h \delta(x, A)$
and so this is an action.

- $\delta \in \mathcal{D}$ is <u>invariant</u> under $G$ if $\tilde{T}_g \delta = \delta$ $\forall g \in G$

<u>Lemma</u> (5) Nonrandomized $\delta$ is invariant iff $d$ is equivariant, i.e. iff $d(T_g x) = T_g^* d(x)$ $\forall x \in \mathcal{X}$, $\forall g \in G$

Proof: So $\delta(x, A) = \begin{cases} 1 & d(x) \in A \\ 0 & \text{otherwise} \end{cases}$, and

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

Taking $A = \{d(x)\}$ this implies that if $\delta$ is invariant, then $d(T_g^{-1}x) = T_g^{*-1}d(x)$ $\forall g$ and so $d(T_g x) = T_g^* d(x)$ $\forall g$. If this holds $\forall x, g$ then $\tilde{T}_g \delta(x, A) = \begin{cases} 1 & d(T_g^{-1}x) \in T_g^{*-1}A \\ 0 & \text{otherwise} \end{cases}$

$$= \begin{cases} 1 & d(x) \in A \\ 0 & \text{otherwise} \end{cases} = \delta(x, A) \text{ and } \delta$$

is invariant.

- let $\mathcal{D}_I \subseteq \mathcal{D}$ denote the class of dec fns invariant under $G$ that leaves the dec problem invariant
- then the (restricted) principle of invariance in decision theory says to look in $\mathcal{D}_I$ for $\delta$ with smallest risk

<u>Lemma</u> ⑥ If $G$ leaves the dec. problem invariant then $R(\bar{T}_g\theta, \tilde{T}_g\delta) = R(\theta, \delta)$.

Proof: $R(\bar{T}_g\theta, \tilde{T}_g\delta)$

$$= \int_{\mathcal{X}} \int_{\mathcal{A}} L(\bar{T}_g\theta, a)\, \tilde{T}_g^* \delta(x, da)\, P_{\bar{T}_g\theta}(dx)$$

$$= \int_{\mathcal{X}} \int_{\mathcal{A}} L(\bar{T}_g\theta, a)\, \delta(T_g^{-1}x, T_g^{*-1}(da))\, P_{\bar{T}_g\theta}(dx)$$

$$\overset{\text{P. 129}}{=} \int_{\mathcal{X}} \int_{\mathcal{A}} L(\bar{T}_g\theta, T_g^* a)\, \delta(T_g^{-1}x, da)\, P_{\bar{T}_g\theta}(dx)$$

$$\overset{\text{invariance of dec problem}}{=} \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a)\, \delta(T_g^{-1}x, da)\, P_\theta \circ T_g^{-1}(dx)$$

$$\overset{\text{P. 129}}{=} \int_{\mathcal{X}} \int_{\mathcal{A}} L(\theta, a)\, \delta(x, da)\, P_\theta(dx) = R(\theta, \delta)$$

<u>Corollary</u> ⑦ If $\delta$ is invariant, then $R(\bar{T}_g\theta, \delta) = R(\theta, \delta)$ namely, the risk fn is constant on orbits on ④

Proof: $R(\bar{T}_g\theta, \delta) = R(\bar{T}_g\theta, \tilde{T}_g\delta) = R(\theta,\delta)$

Corollary (8) If $G$ acts transitively on $\Theta$ and $\delta$ is invariant then $R(\theta,\delta)$ is constant in $\theta$.

- when $G$ acts transitively on $\Theta$ then there will typically be an optimal element of $\mathcal{D}_I$ as real numbers are totally ordered.

- this structure often holds for estimation problems

## (e) Optimal invariant estimators

- suppose the problem is convex so we can restrict to nonrandomized estimators
- suppose $G$ leaves the dec. problem invariant, acts transitively on $\Theta$ and freely on $\mathfrak{X}$
- then pick $\theta_0 \in \Theta$ and for equivariant estimator $d$ the risk is given by

$$R(\theta_0, d) = \int_{\mathfrak{X}} L(\theta_0, d(x))\, P_{\theta_0}(dx)$$

using $x \leftrightarrow ([x], D(x)) = (g, u)$

$D$ is ancillary

$$= \int_{\mathcal{D}} \int_G L(\theta_0, d(T_g u))\, P_{\theta_0 [\cdot]}(dg \mid D)(u)\, P_D(du)$$

$$= \int_{\mathcal{D}} \int_G L(\theta_0, T_g^* d(u))\, P_{\theta_0 [\cdot]}(dg \mid D)(u)\, P_D(du)$$

and if we can find nonrandomized $d$ that minimizes

$$\int_G L(\theta_0, T_g^* d(u))\, P_{\theta_0 [\cdot]}(dg \mid D)(u)$$

for each then $d(x) = T_{[x]}^* d(D(x))$

is the optimal invariant estimator known as the Pitman estimator

## eg location-scale normal

- the model for a sample $\underline{x} = (x_1, \ldots, x_n)'$ is $\{N_n(\mu \underline{1}, \sigma^2 I) : \mu \in \mathbb{R}, \sigma^2 > 0\}$

- the model is invariant under the location-scale group $G = \{(a, c) : a \in \mathbb{R}, c > 0\}$ and the action is transitive on $\Theta$

- a transformation variable is given by $[\underline{x}] = (\bar{x}, s) = (\bar{x}, \|\underline{x} - \bar{x}\underline{1}\|)$ and max invariant $D(\underline{x}) = (\underline{x} - \bar{x}\underline{1})/s$ so $\underline{x} = T_{[\underline{x}]} D(\underline{x}) = \bar{x} + s D(\underline{x})$

- <u>note</u> - $T_{(a_1, c_1)} \underline{x} = T_{(a_2, c_2)} \underline{x}$

  iff $(a_1 - a_2)\underline{1} + (c_1 - c_2)\underline{x} = \underline{0}$

  iff $\underline{x} \in \mathcal{L}\{\underline{1}\}$ and $\underline{x} = \frac{a_1 - a_2}{c_1 - c_2}\underline{1}$ when $c_1 \neq c_2$

  or when $\underline{x} \notin \mathcal{L}\{\underline{1}\}$ then $a_1 = a_2$, $c_1 = c_2$

  - so if we remove $\mathcal{L}\{\underline{1}\}$ (a set of measure 0) from $\mathbb{R}^n$, then the action of $G$ on $\mathbb{R}^n$ is free

- suppose $\Psi(\mu, \sigma^2) = \mu$ and $L(\theta, \psi) = (\theta - \psi)^2/\sigma^2$

- so $\Psi(T_{(a,c)}(\mu, \sigma^2)) = \Psi(a + c\mu, c^2\sigma^2)$ $= a + c\mu$ and so when $\Psi(\mu_1, \sigma_1^2) = \Psi(\mu_2, \sigma_2^2)$

iff $\mu_1 = \mu_2$ then we have $\bar{\Psi}(\bar{T}_{(a,c)}(\mu_1,\sigma_1^2)) = \bar{\Psi}(\bar{T}_{(a,c)}(\mu_2,\sigma_2))$ so $\bar{\Psi}$ is equivariant

and the action on $\bar{\Psi}$ is given by $T^{\Psi}_{(a,c)}\psi = a + c\psi$.

- also $L(\bar{T}_{(a,c)}(\mu,\sigma^2), T^{\Psi}_{(a,c)}\psi)$

$$= ((a+c\mu) - (a+c\psi))^2 / c^2\sigma^2 = L(\mu,\sigma^2)$$

and so the decision problem is invariant under $G$.

- also the problem is convex so we can restrict to nonrandomized estimators based on the mss $(\bar{x}, s)$ and note the mss is equivariant, namely if $(\bar{x}_1, s_1) = (\bar{x}_2, s_2)$ based on $x_1$ and $x_2$ then based on $T_{(a,c)}x_1$, $T_{(a,c)}x_2$ the values of the mss are $(a+c\bar{x}_1, cs_1)$ and $(a+c\bar{x}_2, cs_2)$ which are equal, so the action on the mss is given by $((a+c\bar{x}), cs) = (a,c)(\bar{x}, s)$

- <u>note</u> to reduce to the mss we need to have the mss equivariant so that we can reduce attention to equivariant $d$ that are fns of the mss

- so we can restrict atten to $d(\bar{x}, s)$ satisfying $d(a+c\bar{x}, cs) = a + c\,d(\bar{x}, s)$

- now since $d$ is equivariant (and does not depend on $D(\underset{\sim}{x})$) we have $d(\bar{x},s) = \bar{x} + s\, d(0,1)$ $= \bar{x} + s\, d_0$ and to get the Pitman estimate we need to find the constant $d_0$ minimizing (taking $\Theta_0 = (\mu_0,\sigma_0) = (0,1)$)

$$(*) = \int_{\mathbb{R}\times(0,\infty)} \left(\frac{0-(a+c\,d_0)}{1}\right)^2 P_{(0,1)[\cdot]}\,(d(a,c)\,|\,D)\,(D(\underset{\sim}{x}))$$

- on pages 140-142 we showed that $P_{(0,1)[\cdot]}(\cdot\,|\,D)\,(D(\underset{\sim}{x}))$ is given by (when $\begin{smallmatrix}\mu=0\\ \sigma=1\end{smallmatrix}$)

$$\underset{(f_1)}{a \sim N(0,1/n)} \text{ stat. ind. of } \underset{(f_2)}{s \sim \text{chi}(n-1)}$$

which are both ind. of $D(\underset{\sim}{x})$

- then $(*) = \int_0^\infty \int_{-\infty}^\infty (a+c\,d_0)^2 f_1(a) f_2(c)\,da\,dc$

$$= \int_0^\infty c^2\left(\int_{-\infty}^\infty \left(\frac{a}{c}+d_0\right)^2 f_1(a)\,da\right) f_2(c)\,dc$$

and by least-squares the inner integral is minimized by taking $d_0 = E_{f_1}(a\,|\,c) = E_{f_1}(a)\,|\,c = 0$

- so the Pitman estimator of $\mu$ is given by $d(\bar{x},s) = \bar{x} + s\,d_0 = \bar{x}$ and $\bar{x}$ is the optimal invariant estimator of $\mu$.

(f) Optimal invariant hypothesis tests

- recall we have $\Theta = H_0 \cup H_a$ with $H_0 \cap H_a = \phi$

- $\mathcal{A}(\theta) = 1 - I_{H_0}(\theta)$

- so for $\mathcal{A}$ to be equivariant under $G$ we require if $\mathcal{A}(\theta_1) = \mathcal{A}(\theta_2)$ then $\mathcal{A}(\bar{T}_g\theta_1) = \mathcal{A}(\bar{T}_g\theta_2)$ $\forall g \in G$ which

  occurs iff $\bar{T}_g\theta \in H_0$ whenever $\theta \in H_0$ and $\bar{T}_g H_a \in H_a$ whenever $\theta \in H_a$

- if possible we look for a group s.t. $\bar{T}_g H_0 \subseteq H_0$ and $\bar{T}_g H_a \subseteq H_a$ i.e. is

- when this is the case we have that $T_g^* 1 = 1$ and $T_g^* 0 = 0$ so the action of $G$ on $\mathcal{A}$ is trivial (the identity)

- now $L(\theta, a) = c_0 I_{\{1\}}(a) I_{H_0}(\theta) + c_a I_{\{0\}}(a) I_{H_a}(\theta)$

  and so $L(\bar{T}_g\theta, T_g^* a) = L(\theta, a)$ and the decision problem is invariant under $G$

- $\delta(x, \{1\}) = \varphi(x)$ specifies a dec. fn

- $\delta$ is invariant iff $\delta(T_g x, T_g^*\{1\}) = \delta(T_g x, \{1\}) = \varphi(T_g x)$ $\forall g \in G$ and this implies $\varphi$ is constant on orbits of $G$ on $\mathcal{X}$

and so is a function of the max. invariant $D(x)$

- now suppose $G$ is transitive on $H_0$ so $\bar{T}_g H_0 = H_0$ ($H_0$ is an orbit of $\bar{G}$)

- then $D(x)$ has the same distribution $\forall \theta \in H_0$ using $D(x)$ we are testing a simple hypothesis $H_0$ vs a (usually) composite alternative.

- so we pick $\theta_1 \in H_a$ and find the MP size $\alpha$ test for $H_0$ vs $\theta_1$ based on $D(x)$ and if this doesn't depend on $\theta_1$ then we have found the UMPI (uniformly most powerful invariant) size $\alpha$ test of $H_0$ vs $H_a$.

eg testing the mean: location-scale normal

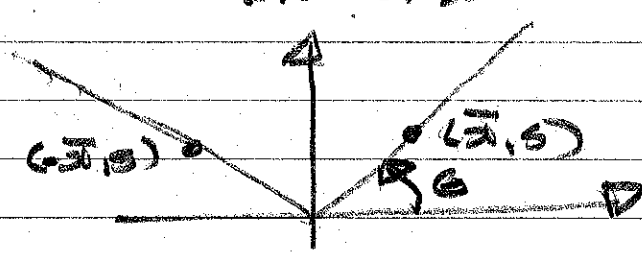
- model for sample $\underline{x} = (x_1, \ldots, x_n)$ is $\{N_n(\mu\underline{1}, \sigma^2 I) : \mu \in R, \sigma^2 > 0\}$
- want to test $H_0 : \mu = 0$ vs $H_a : \mu \neq \mu_0$ so $H_0 = \{0\} \times (0, \infty)$ $\quad H_a = H_0^c$
- we make the reduction to the mss $(\bar{x}, s)$ as there is no loss in power in doing this
- the model for $(\bar{x}, s)$ is for $\mu \in R, \sigma^2 > 0$

$$\bar{x} \sim N(\mu, \sigma^2/n) \text{ ind. of } s \sim \sigma\, chi(n-1)$$

- we want the largest group leaving the decision problem invariant so consider $G = \{(0, c) : c \neq 0\}$ where $(0, c_1)(0, c_2) = (0, c_1 c_2)$
- this acts on $(\bar{x}, s)$ via $T_{(0,c)}(\bar{x}, s) = (c\bar{x}, |c| s)$

Proof: If $T_{(0,c)}(\bar{x}_1, s_1) = (c\bar{x}_1, |c| s_1) = T_{(0,c)}(\bar{x}_2, s_2) = (c\bar{x}_2, |c| s_2)$ so $\bar{x}_1 = \bar{x}_2$, $s_1 = s_2$ and $T_{(0,c)}$ is 1-1. Also $T_{(0,1)}(\bar{x}, s) = (\bar{x}, s)$ while $T_{(0,c_1)(0,c_2)}(\bar{x}, s) = T_{(0,c_1c_2)}(\bar{x}, s) = (c_1 c_2 \bar{x}, |c_1 c_2| s) = T_{(0,c_1)} T_{(0,c_2)}(\bar{x}, s)$.

- the orbit of $(\bar{x}, s)$ under $G$



and these orbits are clearly indexed by
$\cos\theta = |\bar{x}/s|$

- now let $D(\bar{x}, s) = \sqrt{n}\sqrt{n-1}\,|\bar{x}/s|$
serve as the max. invariant

- now note $\bar{x} \sim N(\mu, \sigma^2/n)$ ind of $s \sim \sigma\, \text{chi}(n-1)$

$$t = \frac{\sqrt{n}\,\bar{x}}{\sigma} \Big/ \frac{s}{\sigma\sqrt{n-1}} = \sqrt{n}\sqrt{n-1}\,\frac{\bar{x}}{s}$$

$\sim t_{n-1,\delta}$ = noncentral $t$ on $(n-1)$ df
and noncentrality $\delta$

where $\delta = \sqrt{n}\mu/\sigma$ and this has density

$$g_{n-1,\delta}(t) = \sum_{k=0}^{\infty} \frac{\delta^k}{k!} e^{-\delta^2/2}\, \frac{2^{k/2}\,\Gamma\left(\frac{n+k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{t}{\sqrt{n-1}}\right)^k \left(1+\frac{t^2}{n-1}\right)^{-\frac{n+k}{2}} \frac{1}{\sqrt{n-1}}$$

$$= g_{n-1,0}(t) \sum_{k=0}^{\infty} \frac{\delta^k}{k!} e^{-\frac{\delta^2}{2}}\, 2^{k/2}\frac{\Gamma\left(\frac{n+k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{t}{\sqrt{n-1}}\right)^k \left(1+\frac{t^2}{n-1}\right)^{-\frac{k}{2}}$$

since $g_{n-1,0}(t) = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)} \left(1+\dfrac{t^2}{n-1}\right)^{-\frac{n}{2}} \dfrac{1}{\sqrt{n-1}}$

- <u>note</u> if $t \sim t_{n-1,\delta}$ then $-t \sim t_{n-1,-\delta}$ and so

$$P(|t| \le t_0) = P(t \le t_0) - P(-t \ge t_0)$$

$$= P(t \le t_0) + P(-t \le t_0) - 1$$

and so the density of $|t|$ is given by

$$2\, g_{n-1,0}(t) \sum_{k=0}^{\infty} \frac{\delta^{2k}}{(2k)!}\, 2^k \frac{\Gamma\left(\frac{n+2k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{t^2}{n-1}\right)^k \left(1+\frac{t^2}{n-1}\right)^{-k}$$
$$= h_{n-1,\delta}(|t|)$$

since the terms with odd $k$ cancel and the terms with even $k$ add

- and when $\mu = 0$ so $\delta = 0$ then $|t|$ has density $2 g_{n-1,0}(t) = h_{n-1,0}(|t|)$

- then by the FL the MPI size $\alpha$ test for $\overline{H}_0$ vs $(\mu_1, \sigma) \in H_a$ is to reject when (with $\delta = \sqrt{n}\mu_1/\sigma$)

$$h_{n-1,\delta}(|t|) / h_{n-1,0}(|t|) > c_0 \quad \text{or when}$$

$$\left\{ \sum_{k=0}^{\infty} \frac{\delta^{2k}}{(2k)!}\, 2^k \frac{\Gamma\left(\frac{n+2k}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left( \frac{t^2/n-1}{1+t^2/n-1} \right)^k \right\} > c_0$$

- then $\dfrac{d}{d t^2/n-1} \left(\dfrac{t^2/(n-1)}{1+t^2/n-1}\right)^k = \dfrac{d}{du}\left(\dfrac{u}{1+u}\right)^k$

$$= k \left(\frac{u}{1+u}\right)^{k-1} \left((1-u)^{-1} + u(1-u)^{-2}\right)$$

$$= k \left(\frac{u}{1+u}\right)^{k-1} \left(\frac{1}{1-u}\right)^2 > 0 \quad \text{since } u = \frac{t^2}{n-1} > 0$$

- so differentiating $\{\ \}$ term by term we see that this is an increasing fn of $t^2$ and so the MPI size $\alpha$ test is of the form $a(t) = \begin{cases} 1 & |t| \geq t_{n-1,0,1-\alpha/2} \\ 0 & < \end{cases}$

where $t_{n-1,0,1-\alpha/2}$ is the $1-\alpha/2$ quantile of the $t_{n-1,0}$ distribution

- since this test doesn't involve $(\mu_1, \sigma_1^2)$ we have proven that the $t$-test is the UMPI size $\alpha$ test for $H_0: \mu = 0$ vs $H_a: \mu \neq 0$.

- similarly if we want to test $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$ then putting

$$|t| = \sqrt{n}\sqrt{n-1}\left|\frac{\bar{x}-\mu_0}{s}\right|$$

and $a(t) = 1$ when $|t| \geq t_{n-1,0,1-\alpha}$ is UMPI size $\alpha$ where now the group is $(\mu_0, 1)\, G\, (\mu_0, 1)^{-1} = (\mu_0, 1)\, G\, (-\mu_0, 1)$ $= \{((1-c)\mu_0, c) : c \neq 0\}$

(Proof: $((1-c_1)\mu_0, c_1)((1-c_2)\mu_0, c_2)$

$= ((1-c_1)\mu_0 + c_1(1-c_2)\mu_0, c_1 c_2) = ((1-c_1 c_2)\mu_0, \mu_0))$

with action $((1-c_1)\mu_0, c_1)(\bar{x}, s) = (\mu_0 + c(\bar{x}-\mu_0), cs)$

and max. invariant $\left|\frac{\bar{x}-\mu_0}{s}\right|$

## (a) Invariant confidence regions

- suppose $\Psi : \Theta \xrightarrow{\text{onto}} \Psi$ and $\Psi$ is equivariant under $G$ so we have an action on $\Psi$ given by $T_g^* \psi = \Psi(\bar{T}_g \theta)$ where $\theta$ is s.t. $\Psi(\theta) = \psi$

- let $G_{\psi_0}$ be the subgroup of $G$ that leaves $H_0 : \Psi(\theta) = \psi_0$ vs $H_a : \Psi(\theta) \neq \psi_0$ invariant (the isotropy subgroup at $\psi_0$)

**Lemma** ⓐ $G_{T_g^* \psi_0} = g G_{\psi_0} g^{-1}$

Proof: Let $\theta$ be s.t. $\Psi(\theta) = \psi_0$ and $h \in G_{\psi_0}$. Then

$$T^*_{ghg^{-1}} T_g^* \psi_0 = T^*_{ghg^{-1}} \Psi(\bar{T}_g \theta)$$

$$= T_g^* \Psi(\bar{T}_h \bar{T}_{g^{-1}} \bar{T}_g \theta) = T_g^* \Psi(\bar{T}_h \theta)$$

$$= T_g^* \Psi(\theta) \text{ since } h \in G_{\psi_0} \text{ so } T_h^* \psi_0 = \psi_0$$

This implies $ghg^{-1} \in G_{T_g^* \psi_0}$ $\forall h \in G_{\psi_0}$

and so $g G_{\psi_0} g^{-1} \subseteq G_{T_g^* \psi_0}$. Now let

$h \in G_{T_g^* \psi_0}$. Then $T_h^* T_g^* \psi_0 = T_g^* \psi_0$

and so $\psi_0 = T^*_{g^{-1}} T_h^* T_g^* \psi_0 = T^*_{g^{-1}hg} \psi_0$

which implies $g^{-1}hg \in G_{\psi_0}$ which implies

$h \in g G_{\psi_0} g^{-1}$.

- now suppose $\{\phi_{\psi_0} : \psi_0 \in \Psi\}$ is a class of tests where $\phi_{\psi_0}$ is of size $\alpha$ and is invariant wrt $G_{\psi_0}$ for $H_0 : \Psi(\theta) = \psi_0$ vs $H_a : \Psi(\theta) \neq \psi_0$ and satisfies $\forall \psi_0, g, x$

$$\phi_{T_g^* \psi_0}(T_g x) = \phi_{\psi_0}(x)$$

- put $C(u, x) = \{\psi : \phi_\psi(x) \le u\}$

- then $C(u, T_g x) = \{\psi : \phi_\psi(T_g x) \le u\}$

$$= \{\psi : \phi_{T_g^* T_{g^{-1}}^* \psi}(T_g x) = \phi_{T_{g^{-1}}^* \psi}(x) \le u\}$$

$$= \{T_g^* \psi : \phi_\psi(x) \le u\} = T_g^* \{\psi : \phi_\psi(x) \le u\}$$

$$= T_g^* C(u, x)$$

and $C$ is an equivariant region for $\psi$ and $P(\Psi(\theta) \in C(u,x) \mid \theta) \ge 1-\alpha$ so $C$ is a $(1-\alpha)$ equivariant CR for $\Psi(\theta)$

- if for each $\psi_0 \in \Psi$ the test $\phi_{\psi_0}$ is UMPI size $\alpha$ for $\psi_0$ then $C$ is a UMAI $(1-\alpha)$ CR for $\Psi(\theta)$

- when the tests are nonrandomized $C(x) = \{\psi : \phi_\psi(x) = 0\}$ and $C(T_g x) = \{\psi : \phi_\psi(T_g x) = 0\} = T_g^* C(x)$

eg <u>location - scale normal</u>

- take $G = \{(a,c) : a \in \mathbb{R}, c \neq 0\}$
- then $G_{\mu_0} = \{((1-c)\mu_0, c) : c \neq 0\}$

$$= (\mu_0, 1) G_0 (\mu_0, 1)^{-1}$$

- $\varphi_{\mu_0}(\bar{x}, s) = 1 \quad \sqrt{n}\sqrt{n-1}\left|\frac{\bar{x}-\mu_0}{s}\right| > t_{n-1,0,1-\alpha/2}$

so $C(\bar{x}, s) = \{\mu : \varphi_\mu(\bar{x}, s) = 0\}$

$$= \{\mu : \sqrt{n}\sqrt{n-1}\left|\frac{\bar{x}-\mu_0}{s}\right| < t_{n-1,0,1-\alpha/2}\}$$

$$= \left[\bar{x} \pm \frac{1}{\sqrt{n}} \frac{s}{\sqrt{n-1}} t_{n-1,0,1-\alpha/2}\right]$$

is UMAI $(1-\alpha)$ confidence interval for $\mu$.

- you also go the other way, namely if $C$ is an equivariant $(1-\alpha)$ CR for $\psi$ then define

$$\varphi_{\psi_0}(x) = \int_0^1 I_{\{u : \psi_0 \notin C(u,x)\}}(u)\, du$$

then when $\psi(\theta) = \psi_0$ then

$$E_\theta(\varphi) = P(\psi_0 \notin C(u,x) \mid \theta) \leq 1 - (1-\alpha) = \alpha$$

and $\varphi_{\psi_0}(T_g x) = \int_0^1 I_{\{\psi_0 \notin C(u, T_g x)\}}(u)\, du$

$$= \int_0^1 I_{\{T_g^{*-1}H_0 \,\&\, C(u,x)\}}(u)\, du = a_{T_g^{*-1}H_0}(x)$$

and so $a_{T_g^* H_0}(T_g x) = a_{H_0}(x)$

and the $a_{H_0}$ are invariant under $G_{H_0}$

$\therefore$ a UMAI $(1-\alpha)$ CR leads to
UMPI size tests.