Likelihood Ratio Tests

- References: ① Section 2.3 of Lectures on Algebraic Statistics

② Asymptotic Statistics by van der Vaart (chapter 16)

We will need the following notion of convergence of sets:

" $H_n \longrightarrow H$ " if H is the set of all limits $\lim_{n \to \infty} h_n$ of converging sequences h_n with $h_n \in H_n$ $\forall n \geq 1$ and the limit $h = \lim_{i \to \infty} h_{n_i}$ of converging sequences h_{n_i} with $h_{n_i} \in H_{n_i}$ $\forall i \geq 1$ is such that $h \in H$.

<u>Lemma 0</u>: If $H_n, H \overset{\neq \emptyset}{\subset} \mathbb{R}^k$ such that $H_n \to H$ and $X_n \in \mathbb{R}^k$ is such that $X_n \xrightarrow[D]{} X$ then:

X_n ↳ sequence of random vectors

(converges in distrib.)

$$\text{dist}(X_n, H_n) \xrightarrow[D]{} \text{dist}(X, H) \text{ where}$$

$$\text{dist}(X, H) = \inf\{ d(X, h) : h \in H \}.$$

Set up of Likelihood Ratio Tests

- $X^{(1)}, \dots, X^{(n)}$ are n i.i.d random vectors whose distribution lies in $\mathcal{P}_\Theta = \{ P_\theta : \theta \in \Theta \}$ where $\Theta \subseteq \mathbb{R}^k$.
- Wish to test if true distribution P_θ lies in submodel of $\mathcal{P}_\Theta$ determined by $\Theta_0 \subset \Theta$.

- i.e we wish to test

$H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta \setminus \Theta_0$

- log-likelihood function of P_θ :

$$\ell_n(\theta) = \sum_{i=1}^{n} \log p_\theta(X^{(i)})$$

<u>Defn ①</u> Likelihood ratio statistic for the above hypothesis test is

$$\lambda_n = 2\left(\sup_{\theta \in \Theta} \ell_n(\theta) - \sup_{\theta \in \Theta_0} \ell_n(\theta) \right)$$

- If λ_n is too large, it is unlikely that $\theta_0 \in \Theta_0$ holds.
- Our interest: how is λ_n (asymptotically) distributed? How does this depend on geometry of Θ_0?

Main takeaway: Under H_0, when Θ and Θ_0 are locally linear, λ_n is asymptotically χ^2 (chi-squared) distributed with # degrees of freedom = co-dimension of Θ_0.

- Nice choices of Θ_0:

  affine subspace

  more generally being a smooth manifold

  image of a smooth curve under poly. map.

- We will look at examples where Θ_0 has the above form and see what happens at singularities.

- Set up for examples: P_θ is the normal distribution family

$\{ \mathcal{N}(\theta, Id_{k \times k}) : \theta \in \mathbb{R}^k \}$. (So $\Theta = \mathbb{R}^k$)

Each distribution has density given by

$p_\theta(x) = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}\|x-\theta\|^2\right)$. Then

$$l_n(\theta) = \sum_{i=1}^{n} \log p_\theta(x^{(i)})$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \| x^{(i)} - \theta \|^2 \quad \text{(ignoring constants)}$$

Setting $\overline{x}_n = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$, we get

$$l_n(\theta) = -\frac{n}{2} \| \overline{x}_n - \theta \|^2 - \frac{1}{2} \underbrace{\| x^{(i)} - \overline{x}_n \|^2}_{\text{independent of } \theta}$$

(rearranging terms)

$$\Rightarrow \lambda_n = n \inf_{\theta \in \Theta_0} \| \bar{X}_n - \theta \|^2 = n \inf_{\theta \in \Theta_0} (\bar{X}_n - \theta)^T (\bar{X}_n - \theta)$$

$$\Rightarrow \lambda_n = n \operatorname{dist}(\bar{X}_n, \Theta_0)^2$$

Lemma ② If $\Theta_0 \subset \mathbb{R}^k$ is a d-dimensional linear subspace and $X \sim N(0, \Sigma)$ then

$$f(X) = \inf_{\theta \in \Theta_0} (X - \theta)^T \Sigma^{-1} (X - \theta) \sim \chi^2_{k-d}$$

Pf. $\because$ Σ is positive definite, so is Σ^{-1}

so by Cholesky decomposition $\exists$ invertible C such that $C^T C = \Sigma^{-1}$.

$\because$ $X \sim N(0, \Sigma)$, distribution of $Y = CX$ is given by:

$$\frac{1}{\det C} \cdot \frac{1}{(2\pi)^{k/2}} \frac{\exp\left(-\frac{1}{2}(C^{-1}Y)^T \Sigma^{-1} (C^{-1}Y)\right)}{(\det \Sigma)^{1/2}}$$

$$= \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2} Y^T Y\right)$$

$$\Rightarrow Y \sim \mathcal{N}(0, Id_{k \times k})$$

Also, $$f(X) = \inf_{\theta \in \Theta_0} (X-\theta)^T \Sigma^{-1} (X-\theta)$$

$$= \inf_{\theta \in \Theta_0} (C^{-1}Y-\theta)^T \Sigma^{-1} (C^{-1}Y - \theta)$$

$$= \inf_{\theta \in \Theta_0} (Y - C\theta)^T (Y - C\theta)$$

$$= \inf \|Y - r\|^2$$

$$r \in C\Theta_0$$

$$= \text{dist}(Y, C\Theta_0)^2$$

$C\Theta_0$ is a d-dim. linear subspace, so $\exists$ orthogonal Q such that

$$QC\Theta_0 = \mathbb{R}^d \times \{0\}^{k-d}$$

Then $f(X) = \text{dist}(Y, C\Theta_0)^2$

$$= \text{dist}(QY, QC\Theta_0)^2 \quad (Q \text{ is orthog.})$$

$$= \text{dist}(Z, \mathbb{R}^d \times \{0\}^{k-d})$$

$$= Z_{d+1}^2 + \cdots Z_k^2$$

($\because$ for $i = d+1, \ldots k$ $(Qr)_i = 0$ for $r \in C\Theta_0$

and for $i = 1, \ldots d$, $\exists r$ s.t. $Z_i = Qr_i$)

From above, $Y \sim N(0, I_{k \times k})$ so

$Z = QY \sim \mathcal{N}(0, I_{k\times k})$

$\Rightarrow$ Each component $Z_i \sim N(0,1)$

$\Rightarrow Z_{d+1}^2 + \cdots + Z_k^2 \sim \chi^2_{k-d}$, chi-squared distr. with $k-d$ degrees of freedom. $\square$

- Recall that we assume that for $i = 1, \dots, n$, $X^{(i)}$ are i.i.d with distribution P_{θ_0} for some true parameter $\theta_0 \in \Theta_0$.

- Let λ_{obs} = numerical value of likelihood ratio statistic calculated from some given data set.

- We are interested in the p-value for the likelihood ratio test that is given by : $P_{\theta_0}(\lambda_n \geq \lambda_{obs})$

Example ③ (Θ_0 = d-dim. affine subsp. of $\mathbb{R}^k$)

Recall from before lemma ②, in our example of the normal family $\{\mathcal{N}(\theta, \mathcal{I}_{k\times k}) : \theta \in \mathbb{R}^k\}$ we saw that:

$$\lambda_n = n \inf_{\theta \in \Theta_0} \|\bar{X}_n - \theta\|^2.$$

Rewrite this as:

$$\lambda_n = \inf_{\theta \in \Theta_0} \|\sqrt{n}(\bar{X}_n - \theta_0) - \sqrt{n}(\theta - \theta_0)\|^2$$

Set $Z = \sqrt{n}(\bar{X}_n - \theta_0)$ and $R_n = \sqrt{n}(\Theta_0 - \theta_0)$.

$\because X^{(i)} \sim \mathcal{N}(\theta_0, \mathcal{I}_{k\times k})$ we have that

$Z \sim \mathcal{N}(0, \mathcal{I}_{k\times k})$, and R_n is linear

$$\lambda_n = \inf_{h \in R_n} \|Z-h\|^2 = \inf_{h \in R_n} (Z-h)^T \mathcal{I}_{k\times k} (Z-h)$$

so we can use lemma ② to conclude that $\lambda_n \sim \chi^2_{k-d}$.

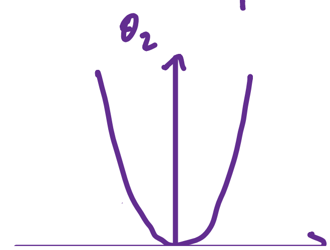
$\Rightarrow$ p-value is

$$P_{\theta_0}(\lambda_n \geq \lambda_{obs}) = P(\chi^2_{k-d} \geq \lambda_{obs}).$$

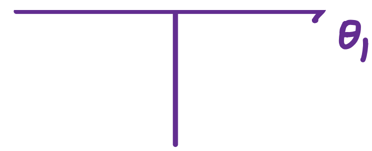
- $R_n = \sqrt{n}(\Theta_0 - \theta_0)$ being a linear subspace was important.
- For Θ_0 = a smooth manifold in $\mathbb{R}^k$, λ_n is asymptotically $\sim \chi^2$.

<u>Example ④</u> (Θ_0 = parabola in the plane)

Specialising the model to



$\{ \mathcal{N}(\theta, I_{2\times 2}) : \theta \in \mathbb{R}^2)$ and



$\Theta_0 = \{ \theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2 : \theta_2 = \theta_1^2 \}$ with

true parameter $\theta_0 = (\theta_{01}, \theta_{02})^T \in \Theta_0$

(so $\theta_{02} = \theta_{01}^2$).

By the same change of variables as

example ③

$$\lambda_n = \inf_{h \in R_n} \| Z - h \|^2 = \operatorname{dist}(Z, R_n) \quad \text{where } Z \sim \mathcal{N}(0, I_{2\times 2}) \qquad R_n \to R$$

where $R_n = \sqrt{n}(\Theta_0 - \theta_0)$

$$= \{ \theta \in \mathbb{R}^2 : \theta_2 = \frac{\theta_1^2}{\sqrt{n}} + 2\theta_{01}\theta_1 \}$$

Can't use lemma ② directly here, but

can do so in the limit.

In the sense of convergence of sets defined in the beginning, $R_n \to R$ where $R = \{\theta \in \mathbb{R}^2 : \theta_2 = 2\theta_{01}\theta_1\}$.

By lemma 0, (with $X_n = X = Z$), it follows that:

$$\lambda_n \xrightarrow{D} \inf_{h \in R} \|Z - h\|^2$$

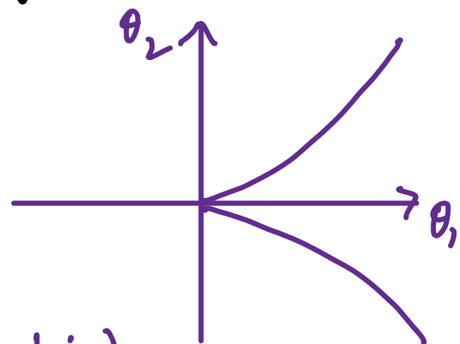
Crucially, R in this case is a 1-dim. linear space (a line in $\mathbb{R}^2$ in the direction $(1, 2\theta_{01})$)

so by lemma 2, the RHS above is χ^2_1 distributed.

$$p = \lim_{n \to \infty} P_{\theta_0}(\lambda_n \geq \lambda_{obs}) = P(\chi^2_1 \geq \lambda_{obs})$$

- Smoothness of Θ_0 was important:
"local geometry of Θ_0, which is captured by linear tangent spaces of fixed dimension, matters for these asymptotics."

- What happens at point of no smoothness? (eg. a corner or cusp)



Example ⑤ (Θ_0 = cuspidal cubic)

Same set up as eqs. ③, ④, i.e $Z \sim N(0, I_{2\times 2})$

but now $\Theta_0 = \{ \theta \in \mathbb{R}^2 : \theta_2^2 = \theta_1^3 \}$.

At $\theta_0 \in \Theta_0 \setminus \{0\}$, curve can be differentiated and using the arguments of example ④

$\lambda_n \xrightarrow{D} \chi_1^2$ for such θ_0.

At $\theta_0 = 0$, $\sqrt{n}\,\Theta_0 = \{ \theta \in \mathbb{R}^2 : \theta_2^{\,2} = \frac{\theta_1^{\,3}}{\sqrt{n}} \}$

and $\sqrt{n}\,\Theta_0 \longrightarrow TC_0(\Theta_0) \longrightarrow$ tangent cone of Θ_0 at 0

$$TC_0(\Theta_0) = \{ \theta \in \mathbb{R}^2 : \theta_2 = 0 \,, \theta_1 \geq 0 \}$$

By lemma 0,

$$\lambda_n \xrightarrow{D} \operatorname{dist}(z, TC_0(\Theta_0))^2$$

But $\operatorname{dist}(z, TC_0(\Theta_0))^2 = \begin{cases} z_2^2 & \text{if } z_1 > 0 \\ z_1^2 + z_2^2 & \text{if } z_1 \leq 0 \end{cases}$

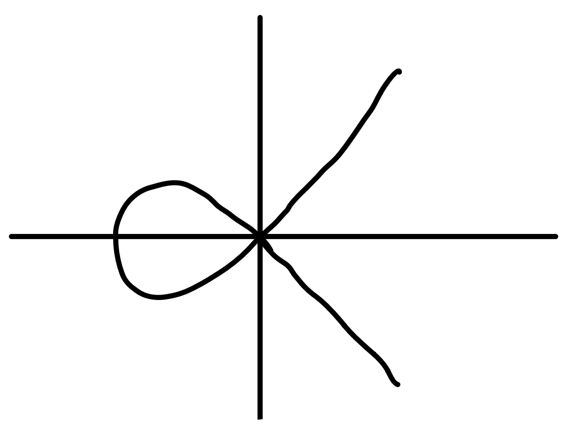
and $\because P(z_1 > 0) = P(z_1 \leq 0) = \frac{1}{2}$,

$\lambda_n \xrightarrow{D} \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ so that the

asymptotic p-value is

$$\frac{1}{2} P(\chi_1^2 \geq \lambda_{obs}) + \frac{1}{2} P(\chi_2^2 \geq \lambda_{obs})$$

- If we erroneously used the p value obtained at non-zero θ_0, how wrong would we be?
- Ans: much smaller than the actual answer above since second term dominates.
- There are also cases where the "naive guess" of the p value is much larger than actual value.
- For example, $\Theta_0 = \{\theta \in \mathbb{R}^2 : \theta_2^2 = \theta_1^2 + \theta_1^3\}$ has a singularity at 0.



- Then $\sqrt{n}\,\Theta_0 = \{\theta \in \mathbb{R}^2 : \theta_2^2 = \theta_1^2 + \frac{\theta_1^3}{\sqrt{n}}\}$

$\Rightarrow TC_0(\Theta_0) = \{\theta \in \mathbb{R}^2 : \theta_2^2 = \theta_1^2\}$

- Intuitively, distance to pair of lines $<$ distance to single line
- $\Rightarrow$ p value $P(\chi_1^2 \geq \lambda_{obs})$ will be much larger than the correct value.

- In example ③, "limiting set" is a 1-dim. linear space and in example ④, it is the positive x-axis.

- The limiting set will always be the tangent cone of Θ_0 at θ_0 $= TC_{\theta_0}(\Theta_0)$
$=$ set of all limits of sequences
$\alpha_n(\theta_n - \theta_0)$ where $\alpha_n \in \mathbb{R}_+$ and
$\theta_n \to \theta_0$ with $\theta_n \in \Theta_0$.

(edit: an earlier version of the notes mistakenly had $\theta_n \in \Theta$ instead of $\theta_n \in \Theta_0$. Thanks, Nils!)

- When θ_0 is not a singularity, $TC_{\theta_0}(\theta_0)$ is the familiar tangent space.

- Properties of $TC_{\theta_0}(\Theta_0)$:

1. It is closed
2. It is a cone.
3. $TC_{\theta_0}(\Theta_1 \cup \Theta_2) = TC_{\theta_0}(\Theta_1) \cup TC_{\theta_0}(\Theta_2)$

provided $\theta_0 \in \Theta_1 \cap \Theta_2$

- The most general Θ_0 we can consider:

Defn. 6 Semi-algebraic set $\Theta_0 \subseteq \mathbb{R}^k$:

$$\Theta_0 = \bigcup_{i=1}^{m} \{ \theta \in \mathbb{R}^k \mid f(\theta) = 0 \text{ for } f \in F_i \text{ and } h(\theta) > 0 \text{ for } h \in H_i \}$$

where $F_i, H_i \subset \mathbb{R}[t_1, \dots t_k]$ are collections of polynomials and all H_i are finite.

"capture how P_θ changes locally with θ"

<u>Defn 7</u> The Fisher information matrix for the model P_Θ, $\Theta \subseteq \mathbb{R}^k$, at $\theta \in \Theta$ is the $k \times k$ matrix with entries

$$\mathcal{I}(\theta)_{ij} = E\left[\left(\frac{\partial}{\partial \theta_i} \log p_\theta(x)\right) \cdot \left(\frac{\partial}{\partial \theta_j} \log p_\theta(x)\right)\right] \qquad 1 \leq i,j \leq k$$

and expectation is taken assuming $X \sim p_\theta$

- $\mathcal{I}(\theta)$ is always positive semi-definite

<u>Example 8</u> (Discrete Fisher information)

Let Θ be the open probability simplex:

$$\Theta = \left\{ \theta \in (0,1)^k : \sum_{i=1}^{k} \theta_i < 1 \right\}$$ and P_Θ be

the corresponding model where each P_θ distribution has density:

$$p_\theta(i) = \theta_i \qquad (\text{for } i = 1, \dots k) \text{ and}$$

$$P_\theta(k+1) = 1 - \sum_{i=1}^{k} \theta_i$$

log-density of P_θ can be written as:

$$\log P_\theta(x) = \sum_{l=1}^{k} \mathbb{1}_{\{x=l\}} \log \theta_l + \mathbb{1}_{\{x=k+1\}} \log(\theta_{k+1})$$

where $\theta_{k+1} = 1 - \sum_{l=1}^{k} \theta_l$

The $$\frac{\partial}{\partial \theta_i} \log p_\theta(x) = \sum_{l=1}^{k} \mathbb{1}_{\{x=l\}} \frac{1}{\theta_l} \delta_{li}$$

$$+ \mathbb{1}_{\{x=k+1\}} \frac{1}{\theta_{k+1}} (-1)$$

$$= \mathbb{1}_{\{x=i\}} \frac{1}{\theta_i} - \mathbb{1}_{\{x=k+1\}} \frac{1}{\theta_{k+1}}$$

$$E\left(\left(\frac{\partial}{\partial \theta_i} \log p_\theta(x)\right)^2\right) = E\left(\mathbb{1}_{\{x=i\}} \frac{1}{\theta_i} - \mathbb{1}_{\{x=k+1\}} \frac{1}{\theta_{k+1}}\right)^2$$

$$= E\left(\mathbb{1}_{\{x=i\}} \frac{1}{\theta_i^2} + \mathbb{1}_{\{x=k+1\}} \frac{1}{\theta_{k+1}^2} - 2 \frac{1}{\theta_i \theta_{k+1}} \mathbb{1}_{\{x=i\}} \mathbb{1}_{\{x=k+1\}} \right)$$

$$= \frac{1}{\theta_i^2} E(\mathbb{1}_{\{x=i\}}) + \frac{1}{\theta_{k+1}^2} E(\mathbb{1}_{\{x=k+1\}}) - 2 \frac{1}{\theta_i \theta_{k+1}} E(\mathbb{1}_{\{x=i\}} \mathbb{1}_{\{x=k+1\}})$$

$$= \frac{1}{\theta_i^2} \cdot \theta_i + \frac{1}{\theta_{k+1}^2} \cdot \theta_{k+1}$$

$$= \frac{1}{\theta_i} + \frac{1}{\theta_{k+1}}$$

$$\Rightarrow \mathcal{J}(\theta)_{ii} = \frac{1}{\theta_i} + \frac{1}{\theta_{k+1}}$$

Similarly $\mathcal{J}(\theta)_{ij} = \frac{1}{\theta_{k+1}}$ for $i \neq j$

If J = all 1's matrix, then

$$\mathcal{I}(\theta) = \frac{1}{\theta_{k+1}} J + D \quad \text{where} \quad D = \operatorname{diag}(\theta_1, \ldots \theta_k)$$

$$= \frac{1}{\theta_{k+1}} \left(u^T u + D_1 \right) \quad D_1 = \theta_{k+1} D.$$

$\Rightarrow \mathcal{I}(\theta)$ is invertible

$$\because \det(u^T u + D') = (1 + u^T D_1^{-1} u) \det D'$$

by the matrix determinant lemma.

<u>Example ⑨</u>

$$\Theta = \mathbb{R}^k, \quad \mathcal{P}_\Theta = \{ \mathcal{N}(0, \mathrm{Id}_{k \times k}) : \theta \in \mathbb{R}^k \}$$

Each p_θ is of the form:

$$p_\theta(x) = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2} \|x - \theta\|^2 \right)$$

$$\Rightarrow \log p_\theta(x) = \log \frac{1}{(2\pi)^{k/2}} - \frac{1}{2} \sum_{l=1}^{k} (x_l - \theta_l)^2$$

$$(2\pi) \qquad \qquad l=1$$

$$\Rightarrow \frac{\partial}{\partial \theta_i} \log p_\theta(x) = -\frac{1}{2} \frac{\partial}{\partial \theta_i} \sum_{l=1}^{k} (x_l - \theta_l)^2$$

$$= \frac{-1}{2} \sum_{l=1}^{k} 2(x_l - \theta_l)(-1)\delta_{li}$$

$$= x_i - \theta_i$$

$$\Rightarrow \mathcal{J}(\theta)_{ij} = E\left((x_i - \theta_i)(x_j - \theta_j)\right).$$

For $i \neq j$,

$$\mathcal{J}(\theta)_{ij} = 0 \quad \because \int_{-\infty}^{\infty} (x_i - \theta_i)\, e^{-\frac{(x_i-\theta_i)^2}{2}} \cdot dx_i = 0$$

(and same for x_j)

$$\mathcal{J}(\theta_{ii}) =$$

$$\frac{1}{(\sqrt{2\pi})^k} \int^{\infty} dx_1\, e^{-\frac{(x_1-\theta_1)^2}{2}} \cdots \int^{\infty} dx_k\, e^{-\frac{(x_k-\theta_k)^2}{2}} \int^{\infty} dx_i\, (x_i-\theta_i)^2 e^{-\frac{(x_i-\theta_i)^2}{2}}$$

$-\infty$ $\qquad$ $-\infty$ $\qquad$ $-\infty$

$$= \frac{1}{(\sqrt{2\pi})^k} (\sqrt{2\pi})^{k-1} \cdot \sqrt{2\pi}\ 2^{-1} \cdot 2.$$

$$= 1.$$

$$\Rightarrow I(\theta) = Id_{k \times k}.$$

<u>Goal</u>: A general result for the asymptotic distribution of λ_n when Θ is semi-algebraic and P_Θ is sufficiently nice

<u>Defn. 10</u> Let P_Θ be a family of prob. distributions on $S \subseteq \mathbb{R}^m$ with densities wrt a measure ν. P_Θ is an <u>exponential family</u> if the following conditions are satisfied: there exists $k \in \mathbb{N}$, a statistic $T: S \to \mathbb{R}^k$ functions $h: \Theta \to \mathbb{R}^k$ and $Z: \Theta \to \mathbb{R}$ such

functions $h: \Theta \to \mathbb{R}$ ...

that each P_θ has ν-density given by

$$p_\theta(x) = \frac{1}{Z(\theta)} \exp\left(h(\theta)^T \cdot T(x) \right), \quad x \in S.$$

Let $\Omega = \left\{ \omega \in \mathbb{R}^k : \int \exp(\omega^T \cdot T(x))\, d\nu(x) < \infty \right\}$

If Ω, Θ are open S and $h: \Theta \to \Omega$ is a diffeomorphism then we say that $\mathcal{P}_\Theta$ is a <u>regular exponential family of order k</u>

<u>Example ⑪</u> (m-variate normal distribution family with $\Sigma = \mathrm{Id}_{m \times m}$)

If $\Theta = \mathbb{R}^m$ and $\mathcal{P}_\Theta = \left\{ N(\theta, \mathrm{Id}_{m\times m}) : \theta \in \mathbb{R}^m \right\}$

Then $p_\theta(x) = \frac{1}{(2\pi)^{m/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}(x_i-\theta_i)^2\right)$

$$= \frac{\exp(-\|\theta\|^2)}{(2\pi)^{m/2}} \exp\left(-\frac{x_1^2}{2} - \dots -\frac{x_m^2}{2} + x_1\theta_1 + \dots + x_m\theta_m\right)$$

$$= \frac{1}{Z(\theta)} \exp\left(h(\theta)^T T(x)\right)$$

where $T : \mathbb{R}^m \to \mathbb{R}^{2m}$ is

$T(x_1, \dots x_m) = \left(-\frac{x_1^2}{2}, \dots -\frac{x_m^2}{2}, x_1, \dots x_m\right)$ and

$h : \Theta \to \mathbb{R}^{2m}$ is

$h(\theta_1, \dots \theta_m) = (1, \dots 1, \theta_1, \dots \theta_m)$

and $Z : \Theta \to \mathbb{R}$ is $Z(\theta) = \left(\frac{\exp(-\|\theta\|^2)}{(2\pi)^{m/2}}\right)^{-1}$

p_θ is reg. exponential of order $2m$.

Example (12) (discrete data)

$$\Theta = \left\{ \theta \in (0,1)^k : \sum_{i=1}^{k} \theta_i < 1 \right\}$$

$$p_\theta(x) = \theta_i \quad \text{if } x = i$$

$$= 1 - \sum_{i=1}^{k} \theta_i \quad \text{if } x = k+1.$$

We can write :

$$p_\theta(x) = \frac{1}{z(\theta)} \exp\left(h(\theta)^T T(x) \right) \quad \text{where}$$

$$z(\theta) = \theta_{k+1}^{-1}$$

$$h(\theta) = \left(\frac{\log \theta_1}{\theta_{k+1}}, \dots \frac{\log \theta_k}{\theta_{k+1}} \right)$$

$$T(i) = e_i \quad \text{if} \quad i = 1, \dots k$$

and $T(k+1) = 0$.

Then $p_\theta(i) = \theta_{k+1} \exp\left(\frac{\log \theta_i}{\theta_{k+1}} \right)$

θ_{k+1}

$$= \theta_{k+1} \cdot \exp\left(\theta \log (\theta_i)^{\frac{1}{\theta_{k+1}}} \right)$$

$$= \theta_{k+1} \cdot \frac{\theta_i}{\theta_{k+1}}$$

$$= \theta_i$$

$$p_\theta(k+1) = \theta_{k+1} \exp(0)$$

$$= \theta_{k+1}$$

$\Rightarrow \mathcal{P}_\Theta = \{ p_\theta : \theta \in \Theta \}$ is regular exponential family of order k.

Properties of reg. exp. families:

1) Parameter space Θ is open

2) $\mathcal{I}(\theta)$ is well-defined and invertible at all $\theta \in \Theta$.

Theorem (13) (Chernoff)

Suppose $\mathcal{P}_\Theta$ is a regular exponential family with parameter space $\Theta \subseteq \mathbb{R}^k$, and let $\Theta_0 \subseteq \Theta$ be semi-algebraic. If $\theta_0 \in \Theta_0$ and $n \to \infty$, then

$$\lambda_n \xrightarrow{D} \inf_{\tau \in TC_{\theta_0}(\Theta_0)} \| Z - \mathcal{I}(\theta_0)^{1/2} \tau \|^2$$

where $Z \sim \mathcal{N}(0, \mathrm{Id}_{k \times k})$

- Most general result so far.
- For $\mathcal{P}_\Theta = \{ \mathcal{N}(\theta, \mathrm{Id}_{k\times k}) : \theta \in \mathbb{R}^k \}$ and Θ_0 = smooth manifold in $\mathbb{R}^k$, $\because \mathcal{I}(\theta) = \mathrm{id}$. (see example (11)) the r.v. on the RHS

above is just

$$\mathrm{dist}(Z, T C_{\theta_0}(\Theta_0))^2 \sim \chi^2_{\text{co-dim}(\Theta_0)},$$ and

we recover the initial results.

- Testing goodness-of-fit between data and 2 submodels:

- i.e. $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1 \setminus \Theta_0$

for two semi-algebraic $\Theta_0 \subset \Theta_1 \subseteq \Theta$.

using

$$\lambda_n = 2\left(\sup_{\theta \in \Theta_1} l_n(\theta) - \sup_{\theta \in \Theta_0} l_n(\theta)\right)$$

which we can write as

$$\lambda_n = 2\left(\sup_{\theta \in \Theta} l_n(\theta) - \sup_{\theta \in \Theta} l_n(\theta)\right) - 2\left(\sup_{\theta \in \Theta_0} l_n(\theta) - \sup_{\theta \in \Theta} l_n(\theta)\right)$$

and in order to get limiting distrib. of λ_n, apply Chernoff to each expression above.

- What does $TC_{\theta_0}(\Theta_0)$ look like?
- If $\Theta_0 = g(\Gamma)$ for some polynomial map $g : \mathbb{R}^d \to \mathbb{R}^k$, where $\Gamma \subseteq \mathbb{R}^d$ is open and semi-algebraic, then:

Proposition (14)

Column space of $J_g(r_0)$ $\subseteq TC_{\theta_0}(\Theta_0)$ where $g(r_0) = \theta_0$.

where $J_g(r) = \left(\frac{\partial g_i}{\partial r_j}(r) \right)_{1 \le i \le k,\ 1 \le j \le d}$

is the Jacobian matrix for parametrisation g.

Pf. Element in col. space of $J_g(r_0)$

$$= \sum_{j=1}^{d} a_j \left(\frac{\partial g_1(r_0)}{\partial r_j}, \frac{\partial g_2(r_0)}{\partial r_j}, \dots \frac{\partial g_k(r_0)}{\partial r_j} \right)$$

$$= \left(\sum_{j=1}^{d} a_j \frac{\partial g_1(r_0)}{\partial r_j}, \dots, \sum_{j=1}^{d} a_j \frac{\partial g_k(r_0)}{\partial r_j} \right)$$

$$= \left(\langle a, \nabla g_1(r_0) \rangle, \dots \langle a, \nabla g_k(r_0) \rangle \right)$$

$(a = (a_1, \dots a_d))$

$$= \|a\|^2 \left(D_{\hat{a}}\, g_1(r_0), \dots D_{\hat{a}}\, g_k(r_0) \right)$$

$= \|a\|^2 \times$ directional derivative of g at r_0 in the direction of a

$\in TC_{\theta_0}(\Theta_0)$

$\square$

- When does equality hold in preceding proposition? dim of the variety obtained by taking Zariski closure of Θ_0

Lemma 15 If $\theta_0 = g(r_0)$ is <u>not</u> a singularity of Θ_0 and rank $J_g(r_0) = \dim \Theta_0$, then $TC_{\theta_0}(\Theta_0) = $ col. span of $J_g(r_0)$

- Singular points of a variety are those at which variety is not locally like a manifold

- To formally define a singularity, we need the vanishing ideal

$$\mathcal{I}(\Theta_0) = \{ f \in \mathbb{R}[t_1, \dots t_k] : f(\theta) = 0 \text{ for all } \theta \in \Theta_0 \}$$

- From alg. geometry, $\mathcal{I}(\Theta_0)$ always has a finite generating set that we can compute, say $f_1, \dots, f_s$.

- $J_f(\theta) = \left(\frac{\partial f_i(\theta)}{\partial t_j} \right)_{1 \le i \le s,\, 1 \le j \le k}$

<u>Defn.</u> 16 A point θ_0 in $\Theta_0 = g(\Gamma)$ is a singularity if rank $J_f(\theta_0) < k - \dim \Theta_0$

egs. cusps, corners.

- Tangent cone at singularity can significantly distort the ordinary χ^2 asymptotics.

- Lemma 15 is inapplicable for singularities so we use $\mathcal{I}(\Theta_0)$ to find a superset of $TC_{\theta_0}(\Theta_0)$, as follows.

- Let θ_0 be a root of $f \in \mathbb{R}[t_1, \dots, t_k]$

· Set $g(t) = f(t+\theta_0)$ and write g as a sum of homogeneous polynomials

$$g(t) = \sum_{h=l}^{L} f_h(t) \quad \text{where } f_h \text{ is homog. of degree } h$$

$$\Rightarrow f(t) = g(t-\theta_0) = \sum_{h=l}^{L} f_h(t-\theta_0)$$

$\because f(\theta_0) = 0$, $l \geq 1$.

Define $f_{\theta_0, \min} = f_l$.

Lemma 17 Let $\theta_0 \in$ semi alg. Θ_0 and consider $f \in \mathbb{R}[t_1, \dots, t_k]$ such that $f(\theta_0) = 0$ and $f(\theta) \geq 0$ $\forall \theta \in \Theta_0$. Then $\forall \tau \in TC_{\theta_0}(\Theta_0)$, $f_{\theta, \min}(\tau) \geq 0$.

Pf. Let $\tau \in TC_{\theta_0}(\theta_0)$ be

$$\tau = \lim_{n\to\infty} \alpha_n (\theta_n - \theta_0)$$ where $\alpha_n > 0$ and $\theta_n \to \theta_0$.

Let $l = \deg f_{\theta_0, min}$.

Then $$f_{\theta_0,min}(\alpha_n(\theta_n-\theta_0)) = \alpha_n^l f_{\theta_0,min}(\theta_n-\theta_0)$$

$$= \alpha_n^l \left(f(\theta_n) - \sum_{h=l+1}^{L} f_h(\theta_n-\theta_0)\right),$$ so that

$$f_{\theta_0,min}(\tau) = \lim_{n\to\infty} f_{\theta_0,min}(\alpha_n(\theta_n-\theta_0))$$

$$= \lim_{n\to\infty} \alpha_n^l \left(f(\theta_n) - \sum_{h=l+1}^{L} f_h(\theta_n-\theta_0)\right)$$

$$= \lim_{n\to\infty} \alpha_n^l f(\theta_n) - \sum_{h=l+1}^{L} \alpha_n^l f_h(\theta_n-\theta_0)$$

≥ 0 $\square$

Applying lemma 17 to each polynomial in

$$\{ f_{\theta_0, min} : f \in \tilde{I}(\theta_0) \} \subseteq \mathbb{R}[t_1, \ldots, t_k]$$

$\downarrow$

tangent cone ideal

- Algebraic variety defined by tangent cone ideal is the algebraic tangent cone, $AC_{\theta_0}(\theta_0)$
- Lemma 17 then says:

$$TC_{\theta_0}(\theta_0) \subseteq AC_{\theta_0}(\theta_0)$$

Example 18

Θ_0 = 3×3 matrices with positive entries, rank ≤ 2 and entries summing to 1.