Three Inference Questions

Decision Theory - what should I do after observing data?

Belief Theory - what do I believe after observing data?

Evidential Analysis - what do data say about hypotheses?

$\Rightarrow$ Evidence is reflected in how much prob is changed by data not what its magnitude is (or was)

<u>Law of Likelihood</u> - If $H_1 \Rightarrow X \sim P(X|H_1)$ and $H_2 \Rightarrow X \sim P(X|H_2)$

Observation $X=x$ is evidence supporting H_1 over H_2 iff

$$P(x|H_1) > P(x|H_2) \Leftrightarrow \frac{P(x|H_1)}{P(x|H_2)} > 1$$

Likelihood ratio $\frac{P(x|H_1)}{P(x|H_2)}$ measures strength of evidence for H_1 over H_2

LR between 1 & 8 - weak evidence

between 8 & 32 - moderate evidence

over 32 - strong evidence

∞ - definitive evidence $\frac{P(\text{data}|H_1)}{P(\text{data}|H_2)}$

<u>Law of Improbability</u> says if prob of observing $X=x$ under H_0 is low, then $X=x$ is evidence against H_0

! specious - lacks relation to second hypothesis

<u>Likelihood Function</u> $\mathcal{L}(\theta|\underline{x}) \propto f(\underline{x}|\theta)$

A model is <u>identifiable</u> if unique parameters imply unique density functions

<u>Kullback-Leibler Divergence</u> - measure of disparity between two distributions

$$KLD(g,f) = E_g\left[\log \frac{g(x)}{f(x)}\right] \geq 0$$

Hellinger Distance provides a lower bound for KLD

$$KLD(g,f) \geq 2[H(f,g)]^2$$

$H_f: X \sim f(x)$ $\quad H_g: X \sim g(x)$ $\quad$ observe $X_1, \dots X_n \overset{iid}{\sim} f$ or g

$$LR_n = \frac{\prod_{i=1}^{n} f(x_i)}{\prod_{i=1}^{n} g(x_i)}$$

How often is LR big when we want it to be small?

$$P_g(LR_n > k) \leq \frac{E_g[LR_n]}{k} = \frac{1}{k} \quad \text{by Markov's}$$

Universal Bound

Universal Bound also holds sequentially

Let $X_1, \dots, X_n \overset{iid}{\sim} g(x)$ $\quad H_g: X \sim g(x)$ $\quad H_f: X \sim f(x)$ $\quad LR_n = \frac{\prod f(x_i)}{\prod g(x_i)}$

As evidence accumulates $LR_n \to 0$

if $X_1, \dots, X_n \overset{iid}{\sim} f(x)$ $LR_n \to \infty$

Likelihood Ratio converges to truth asymptotically

Asymptotic Behavior of LR drives posterior convergences

Maximum Likelihood estimator (MLE)

1. get Likelihood $\mathcal{L}$
2. get log-likelihood $\ell = \log \mathcal{L}$
3. Take derivative $\frac{\partial \ell}{\partial \theta} = \ell'$ (score function)
4. set score function $= 0$
5. solve for parameter

- MLE may not be unique
- MLE may not have analytical solution
- MLEs are often biased
- MLEs are consistent (if conditions met)
  - Identifiable
  - Continuity
  - Compactness
  - Dominance

Invariance of MLE

Let $\hat{\theta}$ be MLE of θ and $\gamma(\theta)$ some one-to-one function of θ

then $\gamma(\hat{\theta})$ is the MLE of $\gamma(\theta)$

- MLE minimizes KLD between truth and model
- MLE is asymptotically efficient (obtains CRLB) + asymptotically normal
- MLE is function of the MSS, but not necessarily the MSS itself

Properties of Estimators

Bias $E[\hat{\theta}-\theta] = b(\hat{\theta})$

Variance $E[(\hat{\theta}-E(\hat{\theta}))^2] = Var(\hat{\theta})$

Mean Square Error (MSE) $E[(\hat{\theta}-\theta)^2] = MSE(\hat{\theta}) = Var(\hat{\theta}) + b^2(\hat{\theta})$

Consistency $\hat{\theta} \to \theta$ as $n \to \infty$

Bayes Estimators

for joint $f(\underline{X}|\theta)$ and prior $f(\theta)$ posterior is

$$f(\theta|\underline{X}) = \frac{f(\underline{x}|\theta)f(\theta)}{\int_\theta f(\underline{x}|\theta)f(\theta)\,d\theta}$$

posterior mean will <u>shrink</u> the sample mean toward prior **mean**

Bayes Estimators trade some bias for reduction in variance

Continuous Mapping Theorem (CMT)

Let $\{X_n\}$ be elements in space S. Let g be function s.t. $g: S \to S'$ with discontinuity points D_g s.t. $P(X \in D_g) = 0$ Then

$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$

$$X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$$

$$X_n \xrightarrow{a.s.} X \Rightarrow g(X_n) \xrightarrow{a.s.} g(X)$$

by CMT $s_n \to \sigma$, but $E[s_n] = E[h(s_n^2)] < hE[s_n^2] = h(\sigma^2) = \sigma$ so biased.

<u>consistent estimators can be biased</u>. Bias converges to 0 (shrinks with sample size)

Inconsistent MLE examples

D. Basu - discontinuous Likelihood Function

Neyman-Scott problem - infinite nuisance parameters (add parameters as increase n)

Score Functions

derivative of log-likelihood

$$s_i = \frac{\partial \ell(\theta; x_i)}{\partial \theta} = \frac{\partial \log f(x_i; \theta)}{\partial \theta}$$

$$S_n = \sum_{i=1}^{n} s_i$$

1. Score function gives MLE
2. Score function is unbiased estimator of 0 $\quad E[S_n] = 0$

   If model wrong, get biased estimate

$$\sqrt{n}(\bar{S}_n) \xrightarrow{d} N(0, Var(S_i)) = N(0, I(\theta))$$

$$\frac{\sqrt{n}(\bar{S}_n - 0)}{\sqrt{Var(S_i)}} \xrightarrow{d} N(0,1)$$

$$Var(S_i) = E[S_i^2] - (\underbrace{E[S_i]}_{0})^2 = E[S_i^2] = E\left[\left(\frac{\partial \log f(x_i)}{\partial \theta}\right)^2\right] = I(\theta)$$

Fisher Information is the variance of the score function

$I_n(\theta) = Var(\sum S_i) = n I(\theta)$ $\qquad$ estimate of information $I_n(\hat{\theta})$

Bartlett's Identity

$Var(S_i) = E[S_i^2] = -E[S_i']$ $\quad$ if the model is correct

Asymptotic Normality of MLE

Taylor expansion $\ell'(\hat{\theta})$ around $\ell_n'(\theta)$ $\qquad \hat{\theta} - \theta \approx \frac{\ell_n'(\theta)}{-\ell_n''(\theta)}$ $\quad \frac{\text{score function}}{\text{information}}$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right)$$

$$\sqrt{n I(\hat{\theta})}\,(\hat{\theta} - \theta) \xrightarrow{d} N(0,1)$$

Robustness

What happens when working model is not the true model?

let g be true model and f be working model

MLE, $\hat{\theta}_n$ converges to θ_g, the value of θ which minimizes the KLD (disparity) between $f(X;\theta)$ and $g(X)$

$$\hat{\theta}_n \to \operatorname{argmin} E_g\left[\log \frac{g(x)}{f(x)}\right] = \theta_g$$

Object of Inference, θ_g

$$\frac{\partial E_g[\log f(x)]}{\partial \theta} := 0 \quad \text{solve for } \theta_g$$

does $\theta_g = E_g[X]$?

If object of inference is not object of interest, find different working model

If yes, consider variance and efficiency

under model failure distribution of MLE is

$$\sqrt{n}(\hat{\theta} - \theta_g) \xrightarrow{d} N\left(0, \frac{b}{a^2}\right) \qquad b = E[(S_i)^2] \qquad a = -E[S_i']$$

estimate b and a with

$$\hat{b} = \frac{1}{n}\sum (S_i|_{\theta=\hat{\theta}})^2 \qquad \hat{a} = \frac{-1}{n}\sum S_i'|_{\theta=\hat{\theta}} \quad \text{to get sandwich estimator } \frac{\hat{b}}{\hat{a}^2}$$

ratio $\frac{\hat{a}}{\hat{b}}$ gauges degree to which Bartlett's 2nd identity fails $\left(= \frac{\text{model based var}}{\text{sample var}}\right)$

Robust adjusted Likelihood

$$\mathcal{L}_R(\theta) = \mathcal{L}(\theta)^{\hat{a}/\hat{b}}$$

under model failure LR for False alternative over θ_g $\frac{\mathcal{L}_n(\theta)}{\mathcal{L}_n(\theta_g)} \to 0$ as $n \to \infty$

$$P\left(\frac{\mathcal{L}(\theta_n)}{\mathcal{L}(\theta_0)} \geq k\right) = \Phi\left[\frac{-\log k}{c^*} - \frac{c^*}{2}\right]$$

for correct model LR + robust adjusted LR

Estimating Equations

Take equation that is function of the data and unknown parameter and solve for unknown parameter

estimating equation unbiased if $E[g(\underline{X};\theta)] = 0$

standardized estimating equation

$$g_s(\underline{X};\theta) = \frac{g(\underline{X};\theta)}{E\left[\frac{\partial g(\underline{X};\theta)}{\partial\theta}\right]}$$

optimal estimating equation has smallest variance in its class

$$\mathrm{var}[g_s(\underline{X};\theta)] = \frac{E[g_s^2]}{\left(E\left[\frac{\partial g}{\partial\theta}\right]\right)^2}$$

true score function is an optimal estimating equation

method of moments

$E[X_i] = \mu \quad \frac{1}{n}\sum X_i - \mu = 0$, solve for μ

$\mathrm{Var}[X_i] = \sigma^2 \quad \frac{1}{n}\sum (X_i-\mu)^2 - \sigma^2 = 0$, solve for σ^2

$$\mathrm{var}[g_s(\underline{X};\theta)] \geq \frac{1}{I_n(\theta)}$$

Lower bound is achieved if function can be written as

$$g(\underline{X};\theta) = a(\theta)\left[T(\underline{X}) - E[T(\underline{X})]\right] = a(\theta)\left[\underbrace{T(\underline{X})}_{\text{function of data alone}} - \underbrace{h(\theta)}_{\text{function of parameter alone}}\right]$$

Cramer-Rao Lower Bound (CRLB) is minimum variance for any **unbiased** estimator of $h(\theta)$

$$\mathrm{Var}(T(\underline{X})) \geq \frac{[h'(\theta)]^2}{I_n(\theta)}$$

Sufficiency

How much can sample be compressed without losing any information?
what is smallest amount of info needed to write down likelihood function

Definition: A statistic $T(\underline{X})$ is a sufficient statistic for $\mathcal{F}$ if the conditional distribution of $\underline{X}$ given $T(\underline{X})$ is same for all distributions in $\mathcal{F}$

$T(\underline{X})$ sufficient if $P(\underline{X} \in A \mid T(\underline{X}) = t; \theta)$ is same for all $\theta \in \Theta$

$T(\underline{X})$ is function of data that has all information about θ

Factorization Theorem $T(\underline{X})$ is sufficient for θ if + only if

$$f_{\underline{X}}(\underline{X};\theta) = g(T(\underline{X});\theta)\, h(\underline{X})$$

where $h(\underline{X})$ does not depend on θ

Minimal sufficiency

- sufficiency is specific to model. If $T(\underline{X})$ sufficient for $\mathcal{F}$, sufficient for $\mathcal{F}' \subseteq \mathcal{F}$
- if $T(\underline{X}) = w(S(\underline{X}))$ and $T(\underline{X})$ sufficient, $S(\underline{X})$ also sufficient
- any 1-to-1 function of a sufficient stat is also a sufficient stat

A sufficient statistic is minimally sufficient if it is a function of every other sufficient stat

The likelihood function is itself the MSS

Find MSS

ratio $\dfrac{f(\underline{x};\theta)}{f(\underline{y};\theta)} = h(\underline{x},\underline{y})$ find stat that will make ratio free of parameters

Rao-Blackwellization

Start with unbiased estimator $\hat{\theta}$ and sufficient statistic $T(\underline{X})$

make new estimator $\tilde{\theta} = E[\hat{\theta} | T(\underline{X})]$

$$Var(\tilde{\theta}) \leq Var(\hat{\theta})$$

Ancillarity

A statistic is ancillary if it contains no information about θ

$S(\underline{X})$ is ancillary if it's distribution doesn't depend on θ

show $f_s(S(x) | \theta)$ doesn't depend on θ

first order ancillary

$E[S(\underline{X}) | \theta]$ doesn't depend on θ

$\Rightarrow$ MSS may not be independent of ancillary statistic

only happens when distribution is complete (Basu's Thm)

Summary

$$\mathcal{L}(\theta; \underline{x}) = f(\underline{x}; \theta)$$

$= g((t,a); \theta) h_1(x)$ by sufficiency [factorization thm]

$= g(t|a; \theta) h_2(a; \theta) h_1(x)$

$= g(t|a; \theta\, h_2(a) h_1(\underline{x})$ by ancillary for θ

$\propto g(t|a; \theta)$

$= g(t; \theta)$ if $T(\underline{X}) \perp A(\underline{X})$ (family is complete)

Completeness

let $f(t|\theta)$ be pdf for statistic $T(\underline{X})$. family of f is complete if

$$E[g(t)] = 0 \ \forall\theta \Rightarrow g(t) = 0 \ \forall\theta$$

Basu's Thm if $T(\underline{X})$ is complete + MSS, then $T(\underline{X})$ is independent of every ancillary stat

Lemma if MSS exists, any CSS is also the MSS

Lehmann-Scheffe Thm if $T(\underline{X})$ is CSS any stat $h(T(\underline{X}))$ w/ finite variance is MVUE of $E[h(T(\underline{X}))]$

condition on CSS to produce estimate w/ smallest variance of all unbiased estimators

Check for completeness

1) exponential families are complete if the interior of the param space is nonempty

$$h(x)c(\theta)\exp\left[\textstyle\sum_j t_j(x)\, w_j(\theta)\right]$$

$$T(\underline{X}) = \left(\sum_{i=1}^n T_j(X_i), \ldots \sum_{i=1}^n T_k(X_i)\right) \text{ is CSS}$$

2) use definition directly

find MVUE

start w/ unbiased estimator (try X_1)

condition on CSS (Blackwellize)

Conditionality, Sufficiency & Likelihood Principles

Conditionality principle - evidence about parameter of interest depends only on the observed data

⇒ Always condition on ancillary statistics

p-values & other probability calculations don't respect conditionality because they depend on sample space (ie. experiments that were not performed)

Sufficiency principle - sufficient statistic carries all the info about the parameter of interest.

⇒ Two data sets with same sufficient stat should yield same evidence about parameter of interest

Likelihood Principle - The likelihood contains all the statistical evidence in the data

⇒ If two likelihoods are the same, the evidence should be the same.

CP + SP ⇒ LP

3 Evidential Quantities

1) strength of evidence - LR

2) propensity for study to yield misleading evidence - reliability of study design $\quad mis_0 = P(LR > k \mid H_0)$, $\quad mis_1 = P(LR < k \mid H_1)$

3) propensity for observed results to be misleading - reliability of data $\quad P(H_0 \mid LR > k)$, $\quad P(H_1 \mid LR < 1/k)$

CP & LP apply to evidence & observed data, not statistical properties of study design

Evidence & operating characteristics are NOT the same thing

Interval Estimation

$X_1, \dots, X_n \sim N(\theta, \sigma^2)$ Estimate θ w/ MVUE or MLE

since $\bar{X} \sim N(\theta, \sigma^2/n)$ $\bar{X}$ will be close to θ

$$P\left(-1.96 \le \frac{\sqrt{n}(\bar{X}-\theta)}{\sigma} \le 1.96\right) = 0.95$$

$\uparrow$ pivot (statistic whose dist. is free of unknown parameters)

Random interval $P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \theta \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 0.95$

Fixed interval $P(\theta - 1.96\frac{\sigma}{\sqrt{n}} \le \bar{X} \le \theta + 1.96\,\sigma/\sqrt{n}) = 0.95$

- $100(1-\alpha)\%$ confidence region $P(I(\underline{X})$ contains $\theta \mid \theta) = P_\theta(\theta \in I(x)) = 1-\alpha \;\; \forall \theta \in \Theta$

  confidence coefficient $1-\alpha$ measures how often procedure captures θ

- Credible interval – Bayesian posterior dist $P(\theta \mid \bar{x})$

  $P(a(\underline{x}) \le \theta \le b(\underline{x}) \mid \underline{X} = \underline{x}) = 1-\alpha$ $\quad [a(\underline{x}), b(\underline{x})]$ is $1-\alpha$ credible int.

- support intervals – A $1/k$ likelihood support interval is

  $\left\{\theta : \frac{L(\theta)}{L(\hat\theta)} \ge 1/k\right\} = \left\{\theta : \frac{L(\hat\theta)}{L(\theta)} \le k\right\}$ where $k > 1$ and $\hat\theta$ is MLE for θ

Criteria for intervals

1) Expected Length, for $\bar{X} \pm 1.96\,\sigma/\sqrt{n}$ $\quad E[2 \cdot 1.96\,\sigma/\sqrt{n}]$

   Pratt Thm – expected length of $C(x) = [L(x), U(x)]$ is sum (integral) of prob. of false coverage over all false values of parameter

   $$E_{\theta^*}(\text{length}(C(x))) = \int_{\theta \ne \theta^*} P_{\theta^*}(\theta \in C(x))\, d\theta$$

2) Unbiasedness – Interval $I(\underline{X})$ is unbiased if

   $P_\theta(I(\underline{X})$ contains $\theta') \le 1-\alpha \quad \forall \theta' \ne \theta$

   where θ' is a false value of θ

   'More likely to contain true value than any false value'

3) Selectivity – Let I_1 & I_2 be two $100(1-\alpha)\%$ CIs

   I_1 is more selective than I_2 if

   $$P_\theta(\theta' \in I_1(\underline{X})) \le P_\theta(\theta' \in I_2(\underline{X})) \quad \forall \theta' \ne \theta$$

   "I_1 tends to exclude false values more often"

Common CIs

1. $X_1, \dots, X_n \sim N(\mu, \sigma^2)$

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \Big/ \sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}} = Z/\sqrt{\chi^2/df} = \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim \underbrace{t^{n-1}}$$

pivot-dist. doesn't depend on μ or σ^2

$$P\left(\bar{X} - t^{n-1}_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t^{n-1}_{\alpha/2} S/\sqrt{n}\right) = 1-\alpha$$

2. $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ $\quad Y_1, \dots, Y_m \sim N(\eta, \sigma^2)$ $\quad X_i \perp Y_j$

$$\bar{X}_n - \bar{Y}_m \sim N\left(\mu-\eta, \sigma^2\left(\tfrac{1}{n}+\tfrac{1}{m}\right)\right)$$

$$\frac{\bar{X}-\bar{Y}-(\mu-\eta)}{\sqrt{\sigma^2(1/n+1/m)}} \Big/ \sqrt{\frac{\frac{(n-1)S_x^2}{\sigma^2} + \frac{(m-1)S_y^2}{\sigma^2}}{n+m-2}} = Z/\sqrt{\chi^2/df} = \frac{\bar{X}-\bar{Y}-(\mu-\eta)}{\sqrt{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t^{n+m-2}$$

$\bar{X}-\bar{Y} \pm t^{n+m-2}_{\alpha/2}\sqrt{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}$ is $100(1-\alpha)\%$ CI for $E[X]-E[Y]$

3. $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ CI for variance

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$P\left(\frac{(n-1)S^2}{a} \le \sigma^2 \le \frac{(n-1)S^2}{b}\right)$ is $100(1-\alpha)\%$ CI for σ^2

could choose a + b to get 1) equal tail areas 2) shortest interval 3) $a=0$, choose b

Robust Large Sample Intervals

for $\hat\theta$ MLE $\quad \sqrt{I_n(\theta)}(\hat\theta-\theta) \xrightarrow{d} N(0,1)$

$\hat\theta \pm z_{\alpha/2}(I(\theta))^{-1/2}$ large sample $100(1-\alpha)\%$ CI

$\hat\theta \pm z_{\alpha/2}(I(\hat\theta))^{-1/2}$ approx. large sample $100(1-\alpha)\%$ CI

also approx if replace $I(\hat\theta)$ with $I_{obs}(\hat\theta) = -\sum \frac{d^2\ell}{d\theta^2}$

neither are consistent ~~est~~ of true info if working model fails

Criteria for CI validity

1. Consistent estimator of parameter $\hat\theta \to \theta$
2. Asymptotic Normality $\sqrt{I_n(\theta)}(\hat\theta-\theta) \xrightarrow{d} N(0,1)$
3. Consistent estimator of information $I_n(\hat\theta)/I_n(\theta) \to 1$

can make intervals robust by using S^2 to estimate variance

$$P\left(\frac{\sqrt{n}|\hat\theta-\theta|}{\sqrt{\sigma^2}}\sqrt{\frac{\sigma^2}{S^2}} \le z_{\alpha/2}\right) \to 1-\alpha$$

1st part $N(0,1)$ by CLT

2nd part $\to 1$ consistent

Hypothesis Testing

Hyp. Testing is choice between H_0 & H_1, type I errors & type II errors used

find critical region that controls type I errors

Neyman-Pearson Lemma if C is critical region such that for H_0 $f(x;\theta_0)$

H_1 $f(x;\theta_1)$

$k>0$

1. $\frac{f_1(x)}{f_0(x)} \geq k \quad \forall \underline{X} \in C$
2. $\frac{f_1(x)}{f_0(x)} < k \quad \forall X \in C^c$
3. $P_0(\underline{X} \in C) = \alpha$

critical region C is most powerful (MP) among all tests of size α or smaller

fix α then maximize power. Possible for LR evidence to disagree w/ test result because Hyp. test changes level of evidence to control α as $n \to \infty$

Significance testing

use p-value as measure of evidence against H_0. (no alternatives, decisions between H_0 + H_1, errors)

$$\text{p-value} = P(\text{data as or more extreme than observed} \mid H_0) = P(T(\underline{X}) \geq T(\underline{x}) \mid H_0)$$

- for composite hypotheses apply N-P lemma to each pairing of ind. hypotheses

if rejection decision rule is free of alternative hypothesis the test is **uniformly most powerful** Note: two-sided tests not UMP

Hypothesis tests are **unbiased** if $P_{\theta_1}(\text{reject } H_0) > P_{\theta_0}(\text{reject } H_0) \ \forall \theta_1 \neq \theta_0$

$$\inf_{\theta \in \Theta_1} 1-\beta(\theta) \geq \alpha$$

Tests are **consistent** if for series $\delta_1, \delta_2, \dots \delta_n$ $1-\beta_{\delta_n}(\theta_1) \to 1$ as $n \to \infty$

Monotone Likelihood Ratio (MLR) family of pmfs or pdfs $g(t|\theta)$ has MLR if

$$\forall \theta_2 > \theta_1 \quad \frac{g(t|\theta_2)}{g(t|\theta_1)} \text{ is monotone function of } t$$

Any exponential family $g(t|\theta) = h(t)c(\theta)\exp[w(\theta)t]$ has MLR if $w(\theta)$ non-decreasing function

Karlin-Rubin Thm - Let $T(X)$ be suff. stat for θ and assume $g(t|\theta)$ has MLR
For any t_0 the test of $H_0: \theta \le \theta_0$ vs. $H_1: \theta > \theta_0$ that rejects ~~when~~ H_0 when

$$T(\underline{X}) > t_0 \text{ is UMP test of size } \alpha = P_{\theta_0}(T(\underline{X}) > t_0)$$

Generalized Likelihood Ratio Test $H_0: \theta \in \Theta_0$ $H_1: \theta \in \Theta_1$ $\Theta_0 \cap \Theta_1 = \emptyset$

$$\delta(\underline{X}) = \begin{cases} 1 & \lambda(\underline{X}) \le \lambda_0 \\ 0 & \text{o.w.} \end{cases}$$

λ_0 chosen s.t. $\alpha = \sup_{\theta \in \Theta_0} P_\theta(\lambda(\underline{X}) \le \lambda_0)$ for

$$\lambda(\underline{x}) = \frac{\sup_{\theta \in \Theta_0} f(\underline{x};\theta)}{\sup_{\theta \in \Theta_1} f(\underline{x};\theta)} = \frac{f(x;\hat{\theta}_0)}{f(x;\hat{\theta}_1)} \doteq \frac{f(x;\hat{\theta}_0)}{f(x;\hat{\theta})}$$

$-2\log\lambda(\underline{X}) \xrightarrow{d} \chi^2_{d-d_0}$ $d = \dim\Theta$ $d_0 = \dim\Theta_0$ under some regularity conditions

Wald Test based on large sample distribution of parameter estimate often MLE + asymptotic normality of MLE. for $H_0: \theta = \theta_0$ v. $H_1: \theta \ne \theta_0$

$$\frac{\hat{\theta} - \theta_0}{\sqrt{Var(\hat{\theta})}} \sim N(0,1)$$

Wald Tests use estimated variance which is consistent under null or alternative

Score Test based on large sample behavior of score function under H_0

$$S_n(\theta_0) = \frac{d\ell(\theta_0)}{d\theta} \quad \text{under } H_0 \quad E_{\theta_0}[S_n(\theta_0)] = 0 \quad Var[S_n(\theta_0)] = I(\theta_0)$$

$$\frac{S_n(\theta_0)}{\sqrt{I_n(\theta_0)}} \sim N(0,1)$$

Score tests use variance under null. If H_0 true good power, o.w. lose power

- GLRT, Wald, and Score are asymptotically equivalent under H_0 but not under H_1
- LRTS invariant to transformations of parameter space
- Wald tests do not work well when parameter is near edge of param space
- Score tests are most powerful for small deviations from H_0

Multiple Testing

when multiple tests performed probability of falsely reject at least one increases with additional tests

one solution is to redefine level of significance based on number of tests

$FWER = P_0(\text{reject at least one } H_0 \text{ falsely}) = 1 - P_0(\text{accept all } H_0) = 1-(1-\alpha)^m$

for m tests with H_0^i v H_1^i size $\alpha_i = \alpha$ $\sum \alpha_i = m\alpha$

Bonferroni correction (very conservative)

compare to $\alpha^{adj} = \alpha/m$ or compare $p^{adj} = mp_i$ to α

- False Discovery Rate

$FDR \simeq E[\text{prop of incorrect rejections}] = P(H_0^i \mid H_0^i \text{ rejected})$ empirical Bayes type estimate

controlling FDR does not control Type I or Type II errors

- Benjamini-Hochberg method control FDR at level γ

find largest i such that $p_{(i)} \le \frac{i\gamma}{m}$ where i is rank of p-value

reject all p-values w/ smaller rank (some more subtle nuances)

Can only control one of per-comparison error rate, FWER or FDR at a time

Bootstrap - use estimate of s.e./CI for sampling dist when analytical solution complex or distribution-free approx is desired

$V_F(T_n) \approx V_{\hat{F}}(T_n) \approx V_{boot}$

$X_1, \dots, X_n \sim F(X)$ $\quad T_n = g(\underline{X})$

1. Estimate $V_F(T_n)$ with $V_{\hat{F}}(T_n)$
2. Approximate $V_{\hat{F}}(T_n)$ using simulation (resampling)

simulation

1. draw resample of size n from $\hat{F}_n(x)$ $\quad X_1^*, \dots X_n^* \sim \hat{F}(X)$
2. calculate $T_n^* = g(\underline{X}^*)$
3. Repeat 1+2 B times
4. Use $T_{n,1}^*, \dots, T_{n,B}^*$ to estimate dist. of T_n

common intervals

1. Normal $\hat{\theta} \pm z_{\alpha/2}\sqrt{V_{boot}}$
2. Percentile $(\hat{\theta}^*_{1-\alpha/2}, \hat{\theta}^*_{\alpha/2})$
3. Pivotal $(2\hat{\theta} - \hat{\theta}^*_{1-\alpha/2}, 2\hat{\theta} - \hat{\theta}^*_{\alpha/2})$
4. Bias-corrected

Bootstrap is asymptotic + $\hat{F}(X)$ must be representative of $F(X)$

Can use double bootstrap to check variance of V_{boot} itself