$\mathcal{M}$ STATISTICAL MODEL → SET OF P. DISTRIBUTIONS ON SAMPLE SPACE X

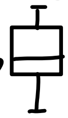
↳ IF X_i IS IID → X = "SAMPLE"

ASSUMPTIONS: X_i IS IID → YOU ONLY NEED THE MARGINAL DISTRIBUTIONS

$L(X, \vartheta) = \prod_{i=1}^{n} f_\vartheta(x_i)$ } MULTIPLY TO GET COMPLETE PMF/PDF

SYMMETRIC, ASYMMETRIC

HISTOGRAM, BOXPLOT



· LOCATION → μ, MEDIAN
· DISPERSION → V, IQR
} THERE ARE ESTIMATORS FOR SAMPLE... AKA FROM THE DATA YOU HAVE

SAMPLE VARIANCE: $S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

QQ PLOT: GRAPH OF DATA, RANKED BY SIZE $\left(F^{-1}\left(\frac{i}{n+1}\right), x_{(i)}\right)$ ← THIS IS THE VALUE YOU EXPECTED FROM THE QUANTILE VS THE QUANTILE YOU OBTAINED

LOCATION-SCALE FAMILY:

$Y = a + bX \quad (b > 0)$

$F_{a,b}(y) = F_X\left(\frac{y-a}{b}\right)$ } YOU CAN ASSOCIATE A FAMILY OF DISTRIBUTIONS TO F

$F^{-1}_{a,b}(\alpha) = a + bF^{-1}(\alpha)$

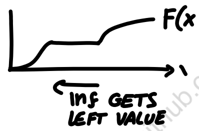
(C^0) $f_Y(y) = f_X\left(\frac{y-a}{b}\right) \cdot \frac{1}{b}$

QUANTILE

$\alpha \mapsto x_\alpha = F^{-1}(\alpha)$ → INVERSE OF CUMULATIVE DISTRIBUTION

$X: F(x) = \alpha$
x IS THE α QUANTILE OF F, FOR $\alpha \in (0,1)$

$\inf\{x : F(x) \geq \alpha\}$



THE NORMAL DISTRIBUTION IS CLOSED UNDER $b > 0$ LINEAR TRANSFORMATIONS

SAMPLE CORRELATION COEFFICIENT $\{(x_i, y_i)\} \mapsto r_{xy}$ → MEASURES LINEAR CORRELATION:
· $r \in [-1, 1]$
· iid $\Rightarrow r_{x,y} \cong 0$
· $r_{x,y} = \pm 1 \Rightarrow y = \bar{y} \pm r \cdot \frac{S_y}{S_x}(x - \bar{x})$

VALIDATE iid:
· X_i NOT CORRELATED WITH X_{i+h} → CHECK AUTOCORR. COEFFICIENT FOR ALL h | $r(h) \approx 0$

{ 1. DEFINE MODEL → $\mathcal{M} = \{P_\vartheta : \vartheta \in \Theta\}$
2. FIND BEST DISTRIBUTION FROM DATA → $\hat{\vartheta}$

ESTIMATOR: $X \mapsto T(X)$ (RANDOM VECTOR) } PRODUCES FROM THE DATA AN ESTIMATE OF SOME $\vartheta \mapsto g(\vartheta)$ PROPERTY

RATING ESTIMATOR PERFORMANCE:

$MSE(\vartheta, T) = \mathbb{E}_\vartheta[\|T - g(\vartheta)\|^2]$

ϑ BEING THE PARAMETERS

P_ϑ DISTRIBUTION, AVERAGED IN ϑ

$= \underbrace{V_\vartheta(T)}_{\text{ERROR}} + \underbrace{(\mathbb{E}_\vartheta[T] - g(\vartheta))^2}_{\text{BIAS}}$

LIKELIHOOD FUNCTION: $\vartheta \mapsto L(\vartheta, x) = \mathbb{P}_\vartheta(x)$

(X iid) $L(\vartheta, x) = \prod_{i=1}^{n} P_\vartheta(x_i)$

} LOG-LIKELIHOOD: FUNCTION $\ell = \log(L)$

UNBIASED ESTIMATOR: BIAS = 0 ⇔ $\mathbb{E}_\vartheta[T] = g(\vartheta)$

SOMETIMES, THE BIASED ESTIMATOR CAN BE MANIPULATED TO GIVE THE UNBIASED ONE

ESTIMATORS GIVING A LINEAR TRANSFORMATION OF A DESIRED QUANTITY, CAN BE MAPPED WITH THE INVERSE RELATION

SO $\mathbb{E}\left(\frac{T-b}{a}\right) = g(\vartheta)$

MLE: PARAMETER $\hat{\vartheta}$ SO $L(\vartheta, x)|_{\vartheta = \hat{\vartheta}}$ IS MAX] + $\hat{\vartheta}$ IS STABLE UNDER COMPOSITION WITH $h(\vartheta)$ FUNCTIONS } WITH MLE USED, TO FIND $h(\vartheta)$, DO $h(\hat{\vartheta})$

$\frac{d}{d\vartheta}\ell(\vartheta, x)$ = SCORE FUNCTION

$= 0$ → FIND $\hat{\vartheta}$ ST. $\begin{cases} \ell'' < 0, \\ \ell' = 0 \end{cases}$

} IF THE FUNCTION IS NOT WELL-BEHAVED, CHECK PATTERNS AND ENDPOINTS

IF THE FUNCTION IS INCREASING IN ϑ, FIND ϑ LARGEST (SUPPORTED BY DATA)

SEE PLUGIN ESTIMATOR FOR FISHER INFORMATION

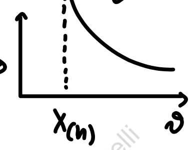
PRONE TO OVERFITTING

$X_i \sim UNIF[0,\vartheta]$

(YOU SOMETIMES DO ALGEBRA ON THE INDICATORS)

$$L(\vartheta; x_1 \ldots x_n) = \prod_{i=1}^{n} P_\vartheta(x_i) = \frac{1}{\vartheta^n} \cdot \prod_{i=1}^{n} I_{[0,\vartheta]}(x) = \frac{1}{\vartheta^n} \cdot I_{[x_{(n)},\infty]}(\vartheta)$$

YOU WANT ϑ AS SMALL AS POSSIBLE



VERY SIMPLE TO IMPLEMENT →

METHOD OF MOMENTS → USE THE FIRST j-TH MOMENTS FROM THE DATA: $\overline{X^j} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$

⊕ COMPARE WITH j-MOMENTS FROM THE PARAMETRISED DISTRIBUTION → FIND ϑ IN TERMS OF MOMENTS

USE THE FIRST k MOMENTS FOR A UNIQUE SOLUTION

MOM CAN BE GENERALISED USING A LOSS FUNCTION: $\min \sum (\text{SAMPLE}_{j\text{-MOM}} - \text{PARAM}_{j\text{-MOM}})^2$ → FOR FITTING MORE MOMENTS THAN YOU HAVE VARIABLES

→ $\mathbb{E}[\bar{x}^2] = \mathbb{E}[\bar{x}]^2 + \frac{1}{n}V(x)$

COMPUTE AS $\propto P_\vartheta(x)\pi(\vartheta)$ (DROP ALL CONST)

INTEGRATE OVER ϑ TO GET NORMALISATION CONSTANT, DIVIDE ⇒ $\frac{P_\vartheta(x)\pi(\vartheta)}{\text{CONST}}$

BAYESIAN APPROACH

RISK: $R(\pi,T) = \int_\Theta \mathbb{E}_\vartheta\left[(T-g(\vartheta))^2\right]\pi(\vartheta)\,d\vartheta$

EVALUATES GOODNESS OF MODEL

YOU WEIGH THE MSE BY ASSOCIATING A PROBABILITY OF ϑ BEING THAT

BAYES ESTIMATOR: $T(x) = \dfrac{\int_\Theta g(\vartheta) P_\vartheta(x)\pi(\vartheta)\,d\vartheta}{\underbrace{\int_\Theta P_\vartheta(x)\,\pi(\vartheta)\,d\vartheta}_{\text{NORMALISING CONSTANT}}}$ → $\vartheta \sim \pi(\vartheta)$ "PRIOR"

$g(\vartheta)$: QUANTITY TO BE ESTIMATED

POSTERIOR DISTRIBUTION: $P_{\bar{\Theta}|X=x}(\vartheta) = \dfrac{P_\vartheta(x)\pi(\vartheta)}{\int_\Theta P_\vartheta(x)\pi(\vartheta)\,d\vartheta}$

IT GENERALLY IS DIFFERENT FROM THE PRIOR $P_\vartheta(x)$

$T(x) = \mathbb{E}\left[g(\vartheta)|X=x\right]$

"CONJUGACY" IF PRIOR → $\pi(\vartheta)$ AND POSTERIOR ARE IN THE SAME FAMILY

POINT ESTIMATE ≠ INTERVAL ESTIMATE

CONFIDENCE REGION → SUBSET OF Θ WITH HIGH $\mathbb{P}$ OF CONTAINING ϑ_{TRUE} } NOT UNIQUE

$P_\vartheta(G_x \ni \vartheta) \geq (1-\alpha)$

AS A CONVENTION, YOU USE UPPER AND LOWER QUANTILES

PIVOT: $T(X,\vartheta)$ WHOSE DISTRIBUTION DOES NOT DEPEND ON ϑ IF $X \sim \vartheta_{TRUE}$ (ASSUMPTION)

$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ } USE IT TO PRODUCE THE CONFIDENCE REGION

NEAR-PIVOT: $T(x,\vartheta)$ WITH DISTRIBUTION THAT CAN BE APPROXIMATED WITHOUT ϑ (USUALLY $n \gg 1$)

CONFIDENCE REGION FOR NORMAL DISTRIBUTION: $\bar{x} \pm \frac{\sigma}{\sqrt{n}}\xi_{1-\alpha/2}$

IF YOU HAVE n SAMPLES USE $\mathcal{L}(\vartheta,x) = \prod p_\vartheta(x)$ →

SCORE FUNCTION: $\frac{\partial}{\partial\vartheta}\log P_\vartheta(x)$

FISHER INFORMATION: $i_\vartheta = \text{Var}_\vartheta[\dot{\ell}_\vartheta(x)] = \mathbb{E}_\vartheta(\dot{\ell}_\vartheta^2(x)) - \underbrace{\mathbb{E}_\vartheta[\dot{\ell}_\vartheta(x)]^2}_{=0}$

ϑ DEPENDENCY

$\dot{\ell}$ IS PARAMETRISED BY ϑ VALUE

CAN BE SEEN A $f(\vartheta)$ TO BE EVALUATED AT ϑ_{TRUE}

/ FOR ϑ TRUE VALUE USE AN ESTIMATE

$i_{\hat{\vartheta}}$ (USING $\hat{\vartheta}_{MLE}$)

PLUG-IN ESTIMATOR: USE $\hat{\vartheta}$ MLE AS YOUR BEST GUESS OF TRUE ϑ

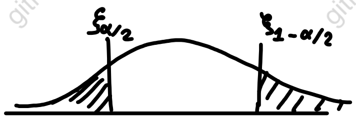
$\hat{i}_\vartheta = -\frac{1}{n}\sum_{i=1}^{n}\ddot{\ell}_\vartheta(x_i)$ } OBSERVED INF. AVERAGE SCORE DERIVATIVE FROM THE DATA

OBSERVED INFORMATION

$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} N(0, \frac{1}{i_\vartheta})$

OR $\sqrt{n i_\vartheta}(\hat{\vartheta}_n - \vartheta) \sim N(0,1)$

} USE IT AS A NEAR PIVOT TO GET CONFIDENCE REGION FOR MLE:



$\vartheta = \hat{\vartheta}_n \pm \frac{1}{\sqrt{n i_\vartheta}} \cdot \xi_{1-\alpha/2}$

$\frac{\alpha}{2}$ IS USED BECAUSE THE DISTRIBUTION IS SYMMETRIC

$\mathcal{M} = \{P_\vartheta : \vartheta \in \Theta\}$

$\Theta = \Theta_0 \cup \Theta_1$ (DISJOINT)

Θ_0: NULL, Θ_1: ALTERNATIVE

- STRONG CONCLUSION → REJECT H_0, ACCEPT H_1
- WEAK CONCLUSION → DO NOT REJECT H_0 (NOT ENOUGH EVIDENCE)

ERROR TYPES

- I → REJECT H_0, THOUGH IT'S CORRECT (BIG NONO)
- II → NOT ACCEPT H_1 THOUGH IT'S CORRECT

TEST: ASSUMING H_0, COMPUTE K

- $x \in K \Rightarrow$ REJECT H_0
- $x \notin K \Rightarrow$ KEEP H_0

STATISTICAL TEST

$K = \{(x_1 \dots x_n) : T(x) \in K_T\}$ → SET OF VALUES OF X, SHOWING H_0 IS FALSE

A SET OF SUFF. UNLIKELY $\{x_i\}$ WILL HAVE ITS CORRESPONDENT $T(x_i)$ VALUES FOR A TEST STATISTIC

K_T: CRITICAL REGION (IN THE SUFF. STATISTICS RANGE)

GENERAL IDEA

$H_0 : \vartheta \in \Theta_0$

$H_1 : \vartheta \in \Theta_1$

A. IF H_0 IS TRUE, $P_\vartheta(X \in K)$ SHOULD BE SMALL

B. IF H_1 IS TRUE, $P_\vartheta(X \notin K)$ SHOULD BE LARGE (YOU WANT TO ACCEPT H_1 IF $\vartheta \in \Theta_1$)

POWER FUNCTION

$\vartheta \mapsto \pi(\vartheta, K) = P_\vartheta(X \in K)$

PICK K SO:

- $\vartheta \in \Theta_0 \mapsto \varepsilon$
- $\vartheta \in \Theta_1 \mapsto 1 - \varepsilon$

$\mathbb{P}$ OF REJECTING H_0 IF $\vartheta_{TRUE} = \vartheta$

$\mathbb{P}$ OF ACCEPTING THE NULL IF $\vartheta_{TRUE} = \vartheta$

SAME IDEA AS CONFIDENCE REGIONS WITH $\mathbb{P}(x \in K) = \alpha$

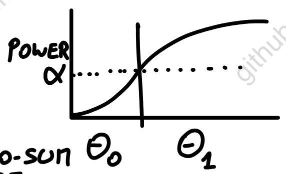
SIZE OF A TEST:

$\alpha_0 = \sup_{\vartheta \in \Theta_0} \pi(\vartheta, K)$

HIGHEST $\mathbb{P}$ ASSIGNED TO Θ_0 WITH K AS THE CRITICAL REGION

HIGHEST $\mathbb{P}$ FOR WHICH TO CONTINUE TO KEEP Θ_0 AS HYPOTHESIS

α_0 IS THE SIGNIFICANCE LEVEL; YOU CAN LOWER IT BUT IT INCREASES $\mathbb{P}$(II ERROR) } ZERO-SUM GAME



IF MULTIPLE TESTS HAVE THE SAME POWER, YOU PICK THE ONE WITH $\pi(\vartheta, K_i)|_{\Theta_1}$ ABOVE ALL THE OTHERS

MORE DATA ALWAYS MEANS MORE POWERFUL TESTS / STEEPER IN Θ_1

YOU HOPE TO FIND POWERFUL TESTS WHICH NEED FEW SAMPLES TO PROVE H_1

	LEFT	RIGHT
ONE SIDED TEST: H_0	$\vartheta \le \vartheta_0$	$\vartheta_0 \le \vartheta$

- TWO SIDED TESTS: $H_0 : \vartheta = \vartheta_0$

→ K_T CRITICAL REGION LOOKS LIKE $\{T \ge c_{\alpha_0}\}$ → IF $T(x)$ "SURPASSES" c_{α_0} OR d_{α_0} (TWO-SIDED TEST) YOU REJECT H_0

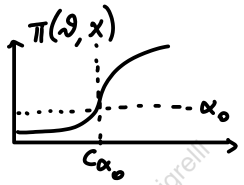
SINCE Θ_0, Θ_1 ARE COMPLEMENTARY, YOU JUST USE A SINGLE SCALAR

FROM α_0, GET $1-\alpha_0$ AS $\mathbb{P}(I) = 1 - \mathbb{P}(II)$ (LOOKING AT TEST POWER)

P-VALUE TESTING

QUANTILE OF POWER FUNCTION

$c_{\alpha_0} = \min\{t : \sup_{\vartheta \in \Theta_0} P_\vartheta(T(x) > t) \le \alpha_0\}$



$p_{VALUE} = \sup_{\vartheta \in \Theta_0} P_\vartheta(T \ge t)$ FOR t → HIGHEST $\mathbb{P}$ FOR $T(x) = T$ TO APPEAR IF Θ_0 IS TRUE

USED IN TESTING:

- $p \le \alpha_0$ → REJECT α_0 (EVIDENCE FOR Θ_1)
- $p > \alpha_0$ → KEEP α_0

EQUIVALENT TO COMPARISON BETWEEN t, c_{α_0}

P-VALUE: $p = 2 \cdot \min\left\{\sup_{\vartheta \in \Theta_0} P_\vartheta(T \le c_{\alpha_0/2}), \sup_{\vartheta \in \Theta_0} P_\vartheta(T \ge c_{1-\alpha_0/2})\right\}$

GAUSS TEST (NORMAL DISTRIBUTION FOR MEAN, KNOWING V)

$X_1 \dots X_n \overset{iid}{\sim} N(\vartheta, \sigma^2)$ | σ^2 KNOWN

$H_0: \vartheta \le \vartheta_0$
$H_1: \vartheta_0 > \vartheta$

IN ϑ TRUE IS LARGER, T WILL HAVE A HIGHER MEAN, WHICH IS WHY THE TEST IS TO THE RIGHT

$T(X) = \sqrt{n}\,\dfrac{\bar{X} - \vartheta_0}{\sigma} \sim N(0,1)$

$K_T = [c_{\alpha_0}, \infty)$

$c_{\alpha_0} = \xi_{1-\alpha_0}$] TEST IS ONE-SIDED

TYPICAL QUESTION:

SIZE OF n SO $\mathbb{P}(\text{II ERROR}) \le \beta$ IF $\vartheta_{TRUE} = \vartheta^*$, $\vartheta^* > \vartheta_0$?

CHANGES $\mathbb{P}$ FROM $\mathbb{P}_{\vartheta_0}$ TO

$\mathbb{P}_{\vartheta^*}(T \notin K_T) = P_{\vartheta^*}(T(X) \le \xi_{1-\alpha_0})$

$= P_{\vartheta^*}\left(\sqrt{n}\,\dfrac{\bar{X}-\vartheta_0}{\sigma} \le \xi_{1-\alpha_0}\right) \le \beta$

$= P_{\vartheta^*}\left(Z \le \xi_{1-\alpha_0} + \dfrac{\vartheta_0}{\sigma/\sqrt{n}} - \dfrac{\vartheta^*}{\sigma/\sqrt{n}}\right)$

$\Phi\left(\xi_{1-\alpha_0} + \dfrac{\vartheta_0}{\sigma/\sqrt{n}} - \dfrac{\vartheta^*}{\sigma/\sqrt{n}}\right) \le \beta$

$\dfrac{\sigma^2}{(\vartheta^*-\vartheta_0)^2}\left(\xi_{1-\alpha_0} - \xi_\beta\right) \le n$

(LEFT-SIDED GAUSS TEST)

$H_0: \vartheta \ge \vartheta_0$
$H_1: \vartheta < \vartheta_0$

$T(X) = \dfrac{\bar{x} - \vartheta_0}{\sigma/\sqrt{n}} \sim N(0,1)$ [IF $\mu = \vartheta_0$]

$K_T = (-\infty, c_{\alpha_0}]$

$c_{\alpha_0} = -\xi_{1-\alpha_0} \longrightarrow t \le \xi_{\alpha_0}$ (REJECT H_0)

$t > \xi_{\alpha_0}$ (RETAIN H_0)

TESTING USING P-VALUES (INVERTS INEQUALITY ORDER)

$t \ge c_{\alpha_0}$
$\Updownarrow$
$\sup_{\vartheta \in \Theta_0} P_\vartheta(T(X) \ge t) \le \alpha_0$

$(c_{\alpha_0} \le t)$ P-VALUE $\le \alpha_0$ $\Rightarrow$ REJECT H_0

$(t < c_{\alpha_0})$ P-VALUE $> \alpha_0$ $\Rightarrow$ KEEP H_0

(TWO-SIDED TEST)

$H_0: \vartheta = \vartheta_0$
$H_1: \vartheta \ne \vartheta_0$

$T(X) = \sqrt{n}\,\dfrac{\bar{x} - \vartheta_0}{\sigma} \sim N(0,1)$ [IF $\mu = \vartheta_0$]

$K_T = (-\infty, c_{\alpha_0}] \cup [c_{\alpha_0}, \infty)$ THANKS TO SYMMETRIC DISTRIBUTION

$c_{\alpha_0} = \xi_{1-\alpha_0/2}$

$t \in K_T \Rightarrow$ REJECT H_0

GAUSS TEST EXAMPLE $\rightarrow 1 - \Phi$ RIGHT-SIDED

P-VALUE $= \sup_{\vartheta \in \Theta_0} P_\vartheta(T(X) \ge t)$

$= P_{\vartheta_0}(T(X) \ge t) = P(Z \ge t)$

$= 1 - \Phi\left(\sqrt{n}\,\dfrac{\bar{x} - \vartheta_0}{\sigma}\right)$

TWO-SIDED:

P-VALUE $= 2 \cdot \min\left\{\sup_{\vartheta \in \Theta_0} P_\vartheta(T(X) \le t),\ \sup_{\vartheta \in \Theta_0} P_\vartheta(T(X) \ge t)\right\}$

GENERAL CASE: P-VALUE = SMALLEST α ST. TEST$_\alpha$ REJECTS H_0

EQUIVALENT METHODS

$t \longleftrightarrow k$] $\rightarrow$ (SEE DATA)

$T(X) \longleftrightarrow K_T$] $\rightarrow [c_{\alpha_0}, \infty)$

$\alpha_0 \longleftrightarrow p$] $\rightarrow \sup_{\vartheta \in \Theta_0} P_\vartheta(T(X) \ge t) \le \alpha_0$

χ-SQUARE

$Z_i \overset{iid}{\sim} N(0,1)$
$W = \sum_i^n Z_i^2$ } n-DEG. FREEDOM

$S = \mathbb{R}$ SUPPORT $\quad \mu = n \quad \sigma^2 = n^2$

STUDENT-t

$\dfrac{Z}{\sqrt{W/n}}$ $\quad \mu = 0 \quad \sigma^2 = \dfrac{n}{n-2}$

IF YOU KNOW σ^2 $\begin{cases} X_1 \dots X_n \sim N(\mu, \sigma^2) \\ \bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right) \end{cases}$ $\rightarrow$ AVERAGING REDUCES VARIANCE

GENERAL ALGORITHM:

- PICK T STATISTIC WITH A KNOWN DISTRIBUTION (USUALLY ESTIMATOR FOR ϑ)
- COMPUTE T DISTRIBUTION ASSUMING Θ_0 $\Rightarrow$ TESTING USING THE POWER FUNCTION

OBVIOUS, SEE μ, σ^2 BEING INDEPENDENT IN VALUE $\longrightarrow$ $\bar{X}, S_X^2$ ARE INDEPENDENT

$(n-1)\dfrac{S_X^2}{\sigma^2} \sim \chi^2_{n-1}$

IF YOU DON'T KNOW V(X) $\longrightarrow$ $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S_X} \sim t_{n-1}$

IN TESTING t-DISTRIBUTION BEHAVES THE SAME AS THE NORMAL (SAME SHAPE)

THE TEST IS MORE POWERFUL BECAUSE ITS DISTRIBUTION HAS MORE DF, SO IT'S MORE CONCENTRATED

TWO-SAMPLE HYPOTHESIS TESTING: $X_1 \dots X_m \overset{iid}{\sim} P_X$, $Y_1 \dots Y_n \overset{iid}{\sim} P_Y$

CHECK IF $P_X \approx P_Y$:

(T-TEST FOR PAIRED) $(X_1, Y_1) \dots (X_n, Y_n) \overset{iid}{\sim} P_{(X,Y)}$

PAIRED USING A DIFFERENT BACKGROUND

$Z_i = X_i - Y_i$ (ASSUMPTION: $Z \sim N(0, \sigma^2)$)

YOU WOULD ASSUME $\Delta\mu = 0$ →

$$\begin{cases} T(Z) = \sqrt{n} \dfrac{\bar{Z}}{S_Z} \sim t_{n-1} \\ V(Z) = V(X) + V(Y) - 2\,cov(X,Y) \end{cases}$$

$H_0 : \mu \geq 0$

$H_1 : \mu < 0$

$$K_T = (-\infty, -c_{\alpha_0}] \overset{!}{=} (-\infty, -t_{n-1, 1-\alpha_0}]$$

(TWO-SAMPLES T-TEST) $X_1 \dots X_m \perp\!\!\!\perp Y_1 \dots Y_n$

$\bar{X} \sim N\left(\mu, \frac{S_X}{m}\right)$ ⊕ NORMALS ARE CLOSED IN LINEAR COMBINATION

$\bar{X} - \bar{Y} \sim N\left(\mu - \nu, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right)$

COMPARE WITH THE ONE FROM THE DATA:

$$T = \frac{\bar{X} - \bar{Y}}{S_{X,Y}\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

H_0 ASSUMPTION IF $\mu = \nu$,

$T \sim t_{n+m-2}$

$$S^2_{XY} = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{n+m-2}$$

S^2_{XY}: χ^2_{n+m-2}; $(m-1)S_X^2$: χ^2_{m-1}; $(n-1)S_Y^2$: χ^2_{n-1}

THE SUM IS ANOTHER χ^2 BECAUSE THE RV ARE $\perp\!\!\!\perp$

$H_0 : \mu - \nu \leq 0$

$K_T = [t_{n+m-2}, \infty)$

(ASYMPTOTIC T-TEST) → FOR $m, n \gg 1$, $S_X \approx S_Y$

$\bar{X} \rightsquigarrow N\left(\mu, \frac{\sigma^2}{m}\right)$

$\bar{Y} \rightsquigarrow N\left(\nu, \frac{\sigma^2}{n}\right)$

$\bar{X} - \bar{Y} \rightsquigarrow N\left(\mu - \nu, \frac{\sigma^2}{m} + \frac{\tau^2}{n}\right)$ } YOU JUST CHECK THE PAIRED MEANS

ASSUMING $\mu = \nu$ ⊕ LCN $(n, m > 20)$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}} \rightsquigarrow N(0,1)$$

$H_0 : \mu \leq \nu$

⇓

$K_T = [c_{\alpha_0}, \infty)$

WILCOXON SIGNED RANK TEST (DISTRIBUTION FREE), $X_1 \dots X_m$, $Y_1 \dots Y_n$

- RANK X BY SIZE IN THE $X \cup Y$ POOL
- $W = \sum_{i=1}^{m} R_i$ → H_0: SAME DISTRIBUTION, X, Y iid

GOODNESS OF FIT

TEST BELONGING TO A FAMILY OF DISTRIBUTIONS

CHI SQUARE: $X_1 \dots X_n \overset{iid}{\sim} F$, F UNKNOWN TEST $H_0 : F = F_0$

(IF $\forall i$ $np_i \geq 5$)

- PARTITION RANGE OF X_i INTO $I_1 \dots I_k$
- $N_j = |\{x_j \mid x_j \in I_j\}|$ ← ELEMENTS THEREIN
- $T = \sum_{j=1}^{k} \frac{(N_j - np_j)^2}{np_j} \overset{H_0}{\sim} \chi^2_{k-1}$
- $K_T = (c_{\alpha_0}, \infty)$

KOLMOGOROV SMIRNOV: $X_1 \dots X_n \overset{iid}{\sim} ?$

F SOME UNKNOWN TEST

$H_0 : F = F_0$

EMPIRICAL DISTRIBUTION FUNCTION $\mathbb{F}_n(x) = \frac{1}{n} \#(x_i \leq x) \overset{!}{=} \frac{1}{n}\sum_{i=1}^{n} I_{x_i \leq x}(x)$

$$T = \sup_x \underbrace{\|\mathbb{F}_n(x) - \mathbb{F}_0(x)\|}$$

FINDS THE GREATEST DEVIATION FROM THE EXPECTED DISTRIBUTION

LIKELIHOOD RATIO TEST (GENERAL METHOD TO FIND A TEST)

↳ $X \sim P_\vartheta$, $H_0 : \vartheta \in \Theta_0$

$$\lambda(x) = \frac{P_{\hat{\vartheta}}(x)}{P_{\hat{\vartheta}_0}(x)}$$ } FOR LARGE VALUES OF $\lambda(x)$ → ACCEPT H_1

$K_T = \{x \mid \lambda(x) \geq c_{\alpha_0}\}$ (SEE $\lambda(x)$ DISTRIBUTION FOR $\vartheta \in \Theta_0$) VARIES BY CASE

TESTS

LOCATION → NORMAL (σ^2 KNOWN)

→ t-TEST ($x \sim N(\mu, \sigma^2)$, μ KNOWN)

→ BINOMIAL (USE THE NORMAL LLN APPROXIMATE)

TWO SAMPLES t-TEST:

- → PAIRED ($\Delta \sim N(0, \sigma^2)$)
- → TWO-SAMPLE ($X \perp\!\!\!\perp Y$, σ^2 SAME, $T = \dfrac{\bar{x} - \bar{y}}{S_{x,y}\sqrt{\frac{1}{m} + \frac{1}{n}}}$)
- → ASYMPTOTIC ($\dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}}$, LARGE m, n)

→ WILCOXON RANK (DISTRIBUTION-LESS)

→ χ^2-TEST (DISCRETISES OBSERVATIONS)

CONFIDENCE REGION FOR THE PARAMETER FOR LEVEL $=\alpha_0$ = CONFIDENCE INTERVAL FOR $\alpha = \alpha_0$ POWER

FIND THE MINIMAL SAMPLE SIZE (TO REJECT THE NULL IF $\mu = \mu_1$)

$$\text{min SAMPLE SIZE} = n : \sup_{\vartheta \in \Theta_1} \mathbb{P}_{\Theta_0}(T(x) \in K_T) \geq \alpha$$

YOU ONLY REJECT H_0 IF $T(x) \in K_T$ BUT DEFAULT $T(X)$ ASSUMES $\vartheta \in \Theta_0$ SO YOU NEED TO CONVERT IT TO Θ_0:

EX $H_0 : \mu \leq \mu_0$ $\qquad T_0(x) = \sqrt{n}\,\dfrac{\bar{x} - \mu_0}{\sigma}$

IF YOU KNOW $\mu = \mu_1$, FIND MIN S. SIZE TO REJECT H_0 WITH $\mathbb{P} = \alpha$ $\qquad T_1(x) = \sqrt{n}\,\dfrac{\bar{x} - \mu_1}{\sigma}$

Θ_0 HAS $K_T = [c_{\alpha_0}, \infty)$ SO

$$\alpha \leq \mathbb{P}_{\Theta_0}\left(\sqrt{n}\,\frac{\bar{x} - \mu_0}{\sigma} \geq c_{\alpha_0}\right)$$

ADD $\sqrt{n}\,\dfrac{\mu_0 - \mu_1}{\sigma}$

$$= \mathbb{P}_{\Theta_1}\left(\sqrt{n}\,\frac{\bar{x} - \mu_1}{\sigma} \geq c_{\alpha_0} + \sqrt{n}\,\frac{\mu_0 - \mu_1}{\sigma}\right)$$

$$= 1 - \Phi\left(c_{\alpha_0} + \sqrt{n}\,\frac{\mu_0 - \mu_1}{\sigma}\right)$$

$$\Phi(\ldots) \leq 1 - \alpha \rightarrow c_{\alpha_0} + \sqrt{n}\,\frac{\mu_0 - \mu_1}{\sigma} \leq c_\alpha$$

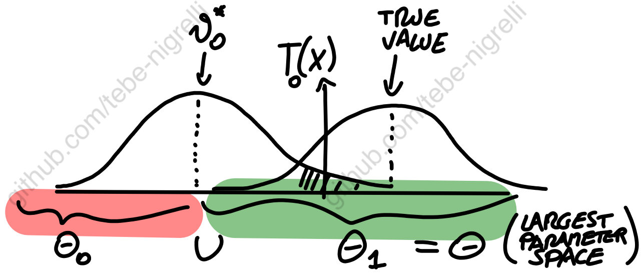
$$n \geq \left(\frac{c_\alpha - c_{\alpha_0}}{\mu_0 - \mu_1}\,\sigma\right)^2$$

$\mu_1 \geq \mu_0$ SO DIVISION BY $\mu_0 - \mu_1$ WILL INVERT THE INEQUALITY

MULTIPLE HYPOTHESIS TESTING: FOR $\begin{cases} \alpha \text{ SIZE} \\ N = \#\text{TESTS} \end{cases}$

$\alpha N = E(\#\text{MISCLASS.})$

→ BONFERRONI CORRECTION: PICK α/N AS SIZE → α MISCLASS. BUT TOO CONSERVATIVE



$T_0(x)$ WILL GIVE A VALUE WHICH IS PREDICTED AS VERY RARE BY Θ_0 BUT COMMON IN Θ_1

CHANGING $T_0(x) \mapsto T_1(x)$ WOULD GIVE A HIGHER PROBABILITY

SINCE $T_0(x)$ IS SO RARE, YOU REJECT H_0, AND SINCE Θ_0 IS COMPLEMENTARY TO Θ_1, IT MEANS YOU ACCEPT Θ_1

OPTIMALITY THEORY → FIND THE BEST ESTIMATOR: REDUCE DATA THEN EXTRACT VALUE

TEBE NIGRELLI 12/04 2024

SUFFICIENT STATISTICS → $V(X)$ WITH ALL RELEVANT INFORMATION

$\Updownarrow$

$\forall x, v$ $P(X=x \mid V=v)$ DOES NOT DEPEND ON ϑ / YOU DON'T NEED TO KNOW ϑ IF YOU HAVE THE DATA AND V

FACTORISATION THEOREM : → USE TO FIND V, OR SHOW IT'S SUFFICIENT

$V=V(X)$ IS SUFF

$\Updownarrow$

$\exists\ v \mapsto g_\vartheta(v)$

$x \mapsto h(x)$ → CAN BE A CONSTANT

ST. $\forall x, \vartheta$

$P_\vartheta(x) = g_\vartheta(V(x))\,h(x)$

UNIQUENESS: → IF $V = f(V^*)$, V^* IS SUFFICIENT → f BIJECTIVE, f PRESERVES "SUFFICIENCY" → IF YOU CAN OBTAIN V FROM V^*, V^* IS SUFF.

OPTIMALITY CRITERIA (DETERMINE THE BEST ESTIMATOR SMALLEST MSE $\forall T, \vartheta$?) IMPOSSIBLE, BECAUSE A CONSTANT HAS ZERO MSE FOR SOME VALUES AND NO ESTIMATOR CAN BE PERFECT ALWAYS

(ALTERNATIVE)

→ BAYES : $\arg\min_T \int MSE(g(\vartheta), T)\,\pi(\vartheta)\,d\vartheta$

→ MINMAX: $\min \sup_{\vartheta\in\Theta} MSE(g(\vartheta), T)$

→ UMVU

IF V IS A FUNCTION OF EVERY OTHER SUFFICIENT STATISTIC ⇒ MINIMAL STATISTIC (SUBSET OF SUFF. STAT.)

MINIMUM VARIANCE UNBIASED ESTIMATOR (UMVU) →
- ESTIMATES $g(\vartheta)$
- UNBIASED
- $V_\vartheta(T) \le V_\vartheta(S)$ $\forall S$ UNBIASED

→ SOMETIMES DOES NOT EXIST

→ BIASED T COULD HAVE LESS VARIANCE

→ NOT PRESERVED BY g MAPS

(SUFF. STAT → UNBIASED ESTIMATOR)

RAO-BLACKWELL → THERE IS AN IMPROVED ESTIMATOR THAT USES THE SUFFICIENT STATISTIC ONLY INSTEAD OF THE WHOLE DATA

$V=V(X)$ SUFF.

$T=T(X)$ ESTIMATOR FOR $g(\vartheta)$

$\Downarrow$

$\exists T^* = T^*(V)$ FOR $g(\vartheta)$

$\begin{cases} E_\vartheta(T^*) = E_\vartheta(T) \\ V_\vartheta(T^*) \le V_\vartheta T \quad \forall\vartheta \text{ PARAM} \end{cases}$

$\Downarrow$

$MSE(g(\vartheta); T^*) \le MSE(g(\vartheta); T)$

COMPLETE STATISTICS → $E_\vartheta f(V) = 0 \quad \forall \vartheta\in\Theta$, $\;\Rightarrow E|f(V)|<\infty$

IS TRUE ONLY IF $P_\vartheta(f(V)=0)=1$

DISPROVE: FIND f ST. ST $E_\vartheta f(V)=0$, f IS NOT CONST. 0

(YOU CAN'T FORM AN UNBIASED ESTIMATOR OF 0 FROM IT)

IF $\exists$ UNBIASED $T(V)$ ESTIMATOR

$\Downarrow$

UNIQUE

FINDING UMVU

FIND $T(V)$ UNBIASED: IF IT IS BIASED UP TO A LINEAR TRANSF, INVERT IT

$E[T] = a\hat{p} + b$

$T^* = \dfrac{T-b}{a}$

YOU LOOK FOR A CONNECTION BETWEEN $\bar{X}$, Σx_i / $V=V(X)$ SUFF. STAT. AND $g(\vartheta)$, LINEARLY COMBINING THEM UNTIL YOU GET THE UNBIASED ESTIMATOR

SINCE $T=T(V)$, $\mathbb{E}[T] = \mathbb{E}[T \mid V]$ AND BY RAO-BLACKWELL THE ESTIMATOR IS UMVU

LEHMANN-SCHEFFÉ

$T=T(V)$ UNBIASED $\oplus$ V SUFF. COMPLETE ⇒ T IS UMVU

(YOU DON'T NEED A MINIMAL STATISTIC)

EXPONENTIAL FAMILY :

$P_\vartheta(x) = c(\vartheta)h(x)e^{\sum_{j=1}^{k} Q_j(\vartheta)V_j(x)}$ (k-DIM FAMILY)

$\Downarrow$

IF $\{(Q_1(\vartheta),\dots,Q_k(\vartheta)) : \vartheta\in\Theta\} \subseteq \mathbb{R}^k$ HAS AN INTERIOR POINT ⇒ V IS SUFFICIENT

$T=T(V)$
- UNBIASED
- USING ONLY $V=(V_1 \dots V_k)$

⇒ T IS UMVU

CRAMER-RAO LOWER BOUND $\left(\text{RELATES } g(\vartheta), I_\vartheta, V(\vartheta)\right)$

$$V_\vartheta(T) \geq \frac{g'(\vartheta)^2}{I_\vartheta} \quad \left(\text{ASSUMING } p_\vartheta \in C^1\right)$$

$\rightarrow g(\vartheta)=\vartheta \Rightarrow V_\vartheta(T) \geq I_\vartheta^{-1}$

$\rightarrow \uparrow I_\vartheta, \downarrow$ BOUND

$\rightarrow$ IF EQUALITY $\Rightarrow T$ IS UMVU

ASYMPTOTIC LOWER BOUND

MLE $\hat\vartheta_n$ $\oplus$ REG. ASSUMPTIONS

$$\sqrt{n}(\hat\vartheta_n - \vartheta) \rightsquigarrow N(0, i_\vartheta^{-1})$$

$$\begin{cases} E_\vartheta \hat\vartheta_n \approx \vartheta \\ V_\vartheta \hat\vartheta_n \approx \frac{i_\vartheta^{-1}}{n} \end{cases} \rightarrow \text{MLE IS UMVU ASYMPTOTICALLY}$$

(NON $\rightarrow$ GENERALLY NOT UMVU)

SUM OF INFORMATION

X, Y INDEP.

$$I_{(X,Y)} = I_X + I_Y$$

TEBE NIGRELLI 2024 12/04

EX:

$X \sim Bin(n, p)$

$$I_p = V_p(\dot{\ell}_p(X)) = V_p\left(\frac{\partial}{\partial p}\log\left(\binom{n}{X}p^X(1-p)^{n-X}\right)\right) = V_p\left(\frac{X-np}{p(1-p)}\right) = \frac{1}{p^2(1-p^2)}V_p(X) = \frac{n}{p(1-p)}$$

$\rightarrow$ CONSTANT

CRAMER-RAO L. BOUND: $\frac{p(1-p)}{n}$

ESTIMATOR:

$\hat p = \frac{X}{n}$

$V_p\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$

SINCE THE BOUND IS SHARP, $\hat p$ IS UMVU

FINDING THE MOST POWERFUL TEST:

$$L(\vartheta_1, \vartheta_0, X) = \frac{p_{\vartheta_1}(X)}{p_{\vartheta_0}(X)}$$

NEYMAN-PEARSON LEMMA: IF FOR c_{α_0}, $P_{\vartheta_0}(L \geq c_{\alpha_0}) = \alpha_0$

$\Downarrow$

$K = \{X : L \geq c_{\alpha_0}\}$ IS THE MOST POWERFUL A α_0 LEVEL

GAUSS TEST $(\overset{H_0}{\mu=\mu_0})$

$L = \ldots$ (NORMALISING COEFFICIENTS SIMPLIFY) $= \exp\left(n\bar X \frac{\mu_1-\mu_0}{\sigma^2} + n\frac{\mu_0^2-\mu_1^2}{2\sigma^2}\right) \sim N(\ldots)$

SEE LEHMANN SCHEFFE THEOREM

UNIFORMLY MOST POWERFUL TEST (JUST MEANS THE TEST IS MORE POWERFUL IN ALL Θ_1)

(POWER FUNCTION $\vartheta \longmapsto \pi(\vartheta, K)$ HAS SIZE α_0 FOR TESTING: $H_0: \vartheta \in \Theta_0$)

IF $\sup_{\vartheta \in \Theta_0} \pi(\vartheta, K) \leq \alpha_0$ AND FOR ALL OTHER K' OF SAME POWER, $\pi(\vartheta : K) \geq \pi(\vartheta, K') \;\; \forall \vartheta \in \Theta_1$

GAUSS TEST (UMP)

$X_i \overset{iid}{\sim} N(\mu, \sigma^2)$ ← KNOWN

$H_0: \mu \leq \mu_0$

$$T = \sqrt{n}\frac{\bar X - \mu_0}{\sigma} > \xi_{1-\alpha_0}$$

UNBIASED TEST: π IS ST. $\forall \vartheta_0, \vartheta_1$: $\pi(\vartheta_0) \leq \alpha_0 \leq \pi(\vartheta_1)$

$(H_0: \vartheta \in \Theta_0)$

$\downarrow$

t-TEST IS UMP AMONG ALL UNBIASED TESTS TO TEST $H_0: \mu \leq \mu_0$

(1D EXPONENTIAL FAMILY)

$V = V(X)$,

$\exists d_{\alpha_0}: P_{\vartheta_0}(V(X) > d_{\alpha_0}) = \alpha_0$

$\Downarrow$

$K = \{x : V(x) \geq d_{\alpha_0}\}$ IS UMP AT α_0 FOR TESTING $H_0: Q(\theta) \leq Q(\vartheta_0)$

(MULTI-DIM. SUFFICIENT STATISTICS)

BERNSTEIN VON-MISES

SINCE THE MODEL IS MORE FLEXIBLE, IT WILL PERFORM BETTER AND

$0 \le SS_{RES} \le SS_{TOT}$

LINEAR REGRESSION (YOU ARE "SMEARING" A GAUSSIAN)

$y_i = \alpha + \beta x_i + e_i \quad \forall i$ DATA

$e_i \overset{iid}{\sim} N(0, \sigma^2)$

$\rightarrow y_i \overset{ind}{\sim} N(\alpha + \beta x_i, \sigma^2)$

SO THEY ARE NOT IDENTICALLY DISTRIBUTED

MLE

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\hat{\beta} = \frac{S_y}{S_x} r_{x,y}$

} UNBIASED

$\hat{\sigma} = \frac{1}{n}\sum^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$

SUM OF SQUARES

$SS_{TOT} = \sum_i (y_i - \bar{y})^2$

$SS_{RES} = \sum_i (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$

$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}}$

TEBE NIGRELLI 2024 12/04

T-TEST, CI FOR β

$H_0: \beta = 0$ (A CONSTANT WOULD BE BETTER)

$T = \frac{\hat{\beta} - \beta_0}{\sqrt{V(\hat{\beta})}}$, $\hat{\beta}$ IS MLE

$(H_0)\ T \sim t_{n-2}$

REJECT IF $|T| \ge t_{n-2, 1-\alpha_0/2}$

CI : $\beta = \hat{\beta} \pm \sqrt{V(\hat{\beta})} \cdot t_{n-2, 1-\alpha_0/2}$

LIKELIHOOD RATIO TEST

DESIGN MATRIX: ONE ROW PER DATA POINT

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$\hat{\beta} = \frac{S_y}{S_x} \cdot r_{x,y} = (X^T X)^{-1} X^T y$

$\hat{\sigma}^2 = SS_{RES}/N = \frac{\|y - X\hat{\beta}\|^2}{n}$

$T = 2\log(\lambda_n) = [\dots] = -n \cdot \log(1 - r_{x,y}^2)$

$T \rightsquigarrow \chi^2_{3-2}$

H_0 REJECTED FOR LARGE $r^2_{x,y}$

MULTIPLE LINEAR REGRESSION (LINEARITY NEEDED ONLY IN PARAMETERS $\rightarrow$ IE USE x_i^2 AS ANOTHER DATA POINT)

$\underline{y} = \underset{\substack{\uparrow \\ \text{DATA MATRIX}}}{X}\underline{\beta} + \underline{e}$, $\underline{e} \sim N(0, \sigma^2 I)$

INCLUDE MODEL INTERCEPT FROM EXTENDING X WITH 1-COLUMN

DUMMY VARIABLE:

USE 0, 1 AS CATEGORICAL VARIABLES } FOR k CLASSES, USE k COLUMNS, ALL ZERO VALUES EXCEPT ONE

IF INTERCEPT IS PRESENT, USE $k-1$ COLUMNS

COMPARING MULTIPLE MODELS $\rightarrow$ USE TESTING TO PICK THE SIMPLER

COCHRAN'S THEOREM

TEST SMALL VS LARGE MODEL

$U \overset{iid}{\sim} \chi_p$, $V \overset{iid}{\sim} \chi_q$ $\rightarrow$ $\underbrace{\frac{U}{p} \cdot \frac{q}{V}}_{\ge 0} \sim F_{p,q}$ DEGREES OF FREEDOM

$y \sim N(\mu, \sigma^2 I)$, $\mu = X\beta$

ARE COVARIATES USEFUL?

$H_0: \beta = 0$

$F = \frac{(SS_{TOT} - SS_{RES})/(p-1)}{SS_{RES}/(n-p)} \overset{H_0}{\sim}$ F DISTRIBUTION $(p-1), (n-p)$

LENGTH OF DATA VECTORS

APPLYING A NONLINEAR TRANSFORMATION TO THE DATA COULD MAKE IT LINEAR

TEST SIGNIFICANCE OF TWO PREDICTORS

t-TEST SINGULARLY CAN FAIL, WHILE THE TWO SHOULD BE INCLUDED IN MODEL

DO: $H_0: \beta_I = \beta_m = 0$ WITH F-TEST

$(1-\alpha)$ CONFIDENCE INTERVAL OF β_j

$\hat{\beta}_j \pm t_{n-p, \alpha/2} \sqrt{(X^T X)^{-1}_{jj}}$

$(1-\alpha)$ CI OF β

$\{\beta : (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) \le \frac{np}{n-p} \hat{\sigma}^2 F_{p, n-p, \alpha}\}$

TESTING ONE PREDICTOR:

$H_0: \beta_j = 0$

$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$

$T_j = \frac{\beta_j - \beta_0}{\hat{\sigma}\sqrt{\frac{n}{n-p}}\sqrt{(X^T X)^{-1}_{jj}}}$

$t_j \sim t_{n-p}$

REJECT H_0 IF $|T_j| \ge t_{n-p, 1-\alpha_0/2}$

$\dim V$

$\dim V_0 \rightarrow$ MODEL WITH FEWER VARIABLES

SEE CL34

VALIDATION

ASSUMPTION ON ERROR: $e_i \sim N(0,\sigma^2)$

→ PLOT $(\hat{y}_i, \hat{e}_i)$
LOOK FOR SAME-HEIGHT BANDS
} SO VARIANCE IS CONSTANT

→ PLOT $(X_{ij}, \hat{e}_i)$
MULTIPLE GROUPS SHOULD HAVE SAME VARIANCE
} USE RATIO TEST AND P-VALUE

→ RESIDUALS HISTOGRAM (SENSITIVE TO BIN CHOICE)

IF VAR IS NOT CONST
⇓
$y \mapsto \log(y)$
$y \mapsto \sqrt{y}$

SHAPIRO-WILK TEST

- CHECKS FOR NORMALITY
- UNRELIABLE FOR SMALL n, TOO SENSITIVE FOR LARGE n

⊕ QQ PLOT OF RESIDUALS TO SHOW NORMALITY

ANOVA (CHECK VARIABLE INFLUENCE ON MODEL)

EXTENSION OF t-TEST TO MULTIPLE GROUPS

$H_0 : \mu_1 = [\dots] = \mu_J$

- MAIN vs INTERACTION
  - MAIN: JUST THE PRESENCE OF SOME VARIABLE
  - INTERACTION: SOME OTHER CATEGORY MATTERS

$$SS_{TOT} = \|(I - P_\mu)y\|^2$$
$$SS_{RES} = \|(I - P_{\mu,\alpha,\beta,\gamma})y\|^2$$
$$SS_\alpha = \|(P_{\mu,\alpha} - P_\mu)y\|^2$$
$$SS_\beta = \|(P_{\mu,\beta} - P_\mu)y\|^2$$
$$SS_{TOT} = SS_\alpha + SS_\beta + SS_\gamma + SS_{RES}$$

CONFUSION MATRIX

TN	FN
FP	TP

- SENSITIVITY → $\frac{TP}{P}$
- SPECIFICITY → $\frac{TN}{N}$

OTHER REGRESSIONS

$$y = g(x) + e$$

MODEL SELECTION (PICK BEST FITTING / SIMPLEST MODEL)

→ OVERFITTING: MODEL IS TOO COMPLEX, CAN'T GENERALISE

BINARY CLASSIFICATION: MODEL $P(Y \mid X = x)$ ⊕ $Y \in \{0,1\}$

$$p(x) = \frac{1}{1 + e^{-\beta \cdot x}}$$

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \underline{\beta} \cdot \underline{x}$$

$x \mapsto \phi(x)$ IS ALSO USED

$\uparrow$ ODDS VALUE

- $\hat{\beta}$ IS FOUND FROM ITERATIVE METHODS: $I_n(\beta)^{1/2}(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} N(0, I_p)$

  $I_n(\hat{\beta})$ USED

- LRT : $2\log\lambda_n \sim \chi^2_{p-p_0}$ $\quad \Big| \quad \lambda_n = \frac{L(\hat{\beta}_1, y)}{L(\hat{\beta}_0, y)}$
- WALD TEST : $(\hat{\beta} - \beta)^T I_n(\beta)(\hat{\beta} - \beta) \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2_p$

PREDICTION: $\mathbb{I}_{(0,\frac{1}{2})}(p(x))$

R^2

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}}$$
$$R^2_{ADJ} = 1 - \frac{n-1}{n-p}(1 - R^2)$$

PENALTY-BASED

$$\max_{(d,\vartheta)} L(\vartheta, X) - Pen(d)$$

$\uparrow$ DIMENSIONS USED BY THE MODEL

AKAIKE INFORMATION CRITERIOR (AIC)

$$\max \log \prod_{i=1}^{n} P_{d,\hat{\vartheta}_d}(x_i) - |d|$$

STEP-UP METHOD:

1. START WITH FULL MODEL
2. TEST ALL $H_0 : \vartheta_i = 0$ SEPARATELY.
   - STOP
   - REJECT ϑ_i WITH HIGHEST P-VALUE

* THERE IS ALSO A STEP DOWN VERSION, WHICH MAY BE DIFFERENT

CROSS-VALIDATION

SPLIT DATA, USE SOME PARTS IN TRAINING, REST FOR EVALUATING

- SPLIT IN 2. TRAIN ON 1 TEST ON 2, DO THE REVERSE, MINIMISE IN d. $SS_{RES_1} + SS_{RES_2}$
- k EQUAL PARTS
  - SPLIT AS n/k
  - VALIDATE ON ONE, TRAIN ON REST

FOR $n \gg 1$ ONLY

PARAMETRIC MODEL → PARAMETER SPACE CAN BE PARAMETRISED BY FINITELY MANY PARAMETERS

↳ HIGH-DIMENSIONAL: # FEATURES >> # SAMPLES } EXTRA ASSUMPTIONS ARE USED → SPARSITY → ONLY USE $s_\beta < n$ VARIABLES

PENALTY TERMS: $\ell_0 \to \hat\beta \in \arg\min \|Y - X\beta\|_2^2 + \lambda s_\beta$

RIDGE → $\hat\beta \in \arg\min \|Y - X\beta\|_2^2 + \lambda_2 \|\beta\|_2^2$

LASSO → $\hat\beta \in \arg\min \|Y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1^2$

ELASTIC NET → $\hat\beta \in \arg\min \|Y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1^2 + \lambda_2 \|\beta\|_2^2$

MATRIX COMPLETION

PAST EXAMS

MLE ||||||| ($+ \hat g(\vartheta)$)

MOM |||

UNBIASED ESTIMATOR CHECK/FIND ||| (SIMPLE LIN. REGR. $Y_1 \dots Y_n, x_1 \dots x_n$ $\mathbb{E}[Y|X=x] = T(x)$ UNBIASED)

BAYES |||||||

FISHER ||||||

WALD INTERVAL +EVALUATE ||||

GAUSS TEST |||||| (STATE ASSUMP.)

OTHER: (PAIR t) (TWO t ||) ($X_{(1)}$ FIND k_T $X_{1 \dots n} \sim e^{-x+\vartheta} I_{x \geq \vartheta}$)

(FIND LRT, k_T: $X_{1 \dots n} \sim 2\vartheta x e^{-\vartheta x^2}$, $H_0: \vartheta = \vartheta_0$) (MIN SAMP. SIZE)

(P-VALUE)

($X_{(1)}$ DISTRIBUTION) ($\overline{X}^2$ DISTR BIAS)

R OUTPUT |||||

(COMPARE LINEAR/LOG REGRESSIONS + FORMULAS) (ANOVA: MODEL FORMULA, ASSUMPTIONS VARIABLE NAMES ||) (DEFINE LOGISTIC, EXPLAIN IN WORDS THE DIFF FROM LINEAR)

FIND SUFF STATISTIC ($X_{1 \dots n} \sim N(\vartheta, \vartheta^2)$ 20) ($X_{1 \dots n} \sim \lambda e^{-\lambda(x-\mu)} I_{x > \mu}$ $\lambda > 0$ μ UNKNOWN)

(IS EXP FAMILY?)

UMVU (BINOM FOR p^2) (SEE SAMPLE 2ND PART #1 EX 2)

CRAMER RAO SHARP ($X_{1 \dots n} \sim N(\mu, \sigma^2)$ T FOR μ) |||||

(NP. LEMMA) (FACT THM |||) (CRAMER RAO |||)