STA 3000Y — Nov. 15, 2013

- likelihood & p-value functions

  model $y \sim f(y;\theta)$ $\quad \theta = (\psi, \lambda)$ $\quad \psi \in \mathbb{R}$

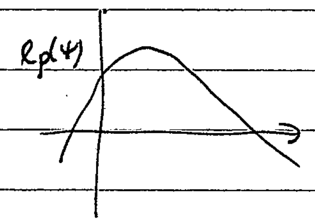
  log-lik. $\ell(\theta; y)$

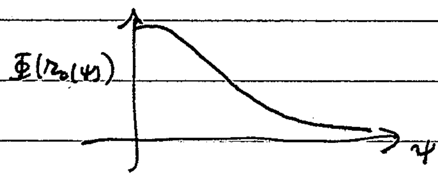
  profile log-lik $\ell_p(\psi;y) = \ell(\psi, \hat\lambda_\psi; y)$ $\qquad \frac{\partial \ell(\theta)}{\partial\lambda}\Big|_{\hat\lambda_\psi} = 0$

  approx. pivot $w(\theta) = 2\{\ell(\hat\theta;y) - \ell(\theta;y)\} \sim \chi^2_p$

$$r(\psi) = \pm\sqrt{2\{\ell_p(\hat\psi;y) - \ell_p(\psi;y)\}} \;\dot\sim\; N(0,1)$$

$$P_{\psi,\lambda}(r(\psi) \le r_0(\psi)) \doteq \Pr\{Z \le r_0(\psi)\} = \Phi(r_0(\psi))$$





- Notes
  - if $\theta \in \mathbb{R}$ then $\uparrow$ applies w $\ell(\theta)$ & $\pm\sqrt{w(\theta)}$
  - if dim λ large then $\dot\sim N$ will be poor, & if dim $\to\infty$ $\sim N$ may be just wrong
  - in special cases we can find $f_c(s|t;\psi)$ or $f_m(t;\psi)$
    i.e. we can find statistics $s = s(y)$, $t = t(y)$ s.t.
    a) (s,t) are suff't
    b) λ is eliminated exactly by cond'g or marg'g
  - in these case $\ell_c(\psi; s|t)$ or $\ell_m(\psi; t)$ can be used ~~as above~~ instead of ℓ_p, and $\sim N$ will generally be better than for profile, even if dim λ large
  - if $\psi \in \mathbb{R}^q$ then same arguments apply, but plots are more complicated, ~~because~~
  - equivalent to plotting $\Phi(r_0(\psi))$ vs ψ is plot of $r_0(\psi)$ vs ψ
  - $\exists$ other asy. equiv. pivots eg $(\hat\psi - \psi)\, j_p^{1/2}(\hat\psi)$ etc.

- classical est^n & testing

Pure significance tests (CH Ch. 3, SM 7.3.1) Princ. Ch 3

Instead of a family of models for data $\{f(y;\theta);\theta\in\Theta\}$, suppose we have only a (null) hypothesis H_0 which specifies the dist of Y or of some summary $T=t(Y)$.

CH (4) e.g. • H_0 corresponds to prediction of scientific theory — thought to be true
• may divide dist's into qual. different types, but these are not dist. of data consist w H_0 — may not be true

{ • simple set of circumstances e.g. data ~ N or data ~ iid
• absence of structure e.g. data ~ PP or TS ~~iid~~ has no corr }
g.o.f.t

RC (2) 2: - subj.-matter hyp., may be expected to be true
· adequacy of a formal model is to be assessed — g o f t

Princ (5)
- may be true — $H_0: \psi=\psi_0$
- not true, but divides par. into regions w different interp.
- may permit simp. of a model for interp.
- only ψ_0 model is feasible but this is to provide a basis for assessing dep. ~~qual dep.~~
- defined but some idea of dep. of interest

] g o f t

examples : • sequence of obs'd events from LHC is background only

$\cancel{Y_1,\ldots,Y_n}$ iid $Po(b)$ $(Y\geq y_{obs})$

$\cancel{S=\Sigma Y_i}\sim Po(\cancel{n}b)$ $Pr_{H_0}(S\geq s_{obs})$

$=\cancel{\sum_{s_{obs}}^{\infty}\frac{b^s e^{-b}}{s!}}$

• temp. record is consistent w background variation

• treat A = treat B

• $E(Y_i\mid x_i)=\beta_0+\beta_1x_i+\beta_2x_i^2$ $H_0:\beta_2=0$

Assume we have such an H_0, & a st. $T = t(Y)$, s.t. large values of T are evidence against H_0.

p_{obs} = obs'd level of significance
= p-value for testing H_0

$\mathrel{\hat=} \Pr_{H_0}(T \geq t_{obs})$ $\qquad p_0 = \Pr_{H_0}(T \geq t_0)$ $\quad t^\circ$

Small p_{obs} suggests data y_{obs} (via $t(y) = t_{obs}$) is not consistent w H_0

- e.g. $b = 6.7$ $\quad y_{obs} = 23$ $\quad p_{obs} = \sum_{y=23}^{\infty} \frac{(6.7)^y e^{-6.7}}{y!}$

- eg CH 3.1 $\quad Y_i$ are angles between $(0, 2\pi)$ $\quad Y_1, \dots, Y_n$ iid

H_0 $\quad Y_i \sim U(0, 2\pi)$ $\qquad T = t(y) = \sum_{i=1}^{n} \cos Y_i$

$$E_{H_0}(e^{\phi T}) = E(e^{\phi \sum \cos Y_j}) = \prod_{j=1}^{n} E(e^{\phi \cos Y_j}) = \{M_{\cos Y_j}(\phi)\}^n$$

$$M_{\cos Y_j}(\phi) = \frac{1}{2\pi}\int_0^{2\pi} e^{\phi \cos y}\, dy = \{I_0(\phi)\}^n, \text{ say}$$

$\therefore$ density of T available by inversion of mgf.
or, more simply $E_{H_0} T = 0$, $\mathrm{var}_{H_0}(T) = \frac{n}{2}$ $\qquad T \mathrel{\dot\sim} N(0, \frac{n}{2})$ under H_0

$$p_{obs} \doteq \Pr\{N(0, \tfrac{n}{2}) \geq t_{obs} = \sum_{i=1}^{n} (\cos y_i)\}$$

P.I.T. $\qquad p_{obs} = 1 - F_0(t_{obs})$ $\qquad \sim U(0,1)$ under H_0

$$P = 1 - F_0(T) \qquad \Pr_{H_0}(P \leq u) = \Pr_{H_0}\{F_0^{-1}(1-u) \leq T\}$$
$$= 1 - F_0(F_0^{-1}(1-u)) = u$$

CH – "if we were to regard obs data as just decisive against H_0, we would be mistaken in prob p_{obs}"

≠ "prob H_0 is true"

– problems in p-values
- may be very small for meaningless discrepancies
- may be not small for meaningful discrepancies
- don't give info on what if H_0 is not true
- often misinterpreted
- often used as arbitrary cut-offs for 'sig', 'n.s.' usually $p = 0.05$ or 0.01

<u>Example</u> $Y_1, \ldots, Y_n$ iid $F(\cdot)$ (H_0)

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{Y_i \le t\}$$

$$E\hat{F}(t) = F(t)$$

$$\text{var}\,\hat{F}(t) = \frac{1}{n} F(t)\{1-F(t)\}$$

$$E\{\hat{F}(\cdot);\, t \in \mathbb{R}\} = F(\cdot) \quad \text{etc.}$$

popular $T = t(Y)$:

$$\sup_t |\hat{F}_n(t) - F(t)| \qquad \text{K-S}$$

$$\int \{\hat{F}(t) - F(t)\}^2 \, dF(t) \qquad \text{C von M}$$

$$\int \frac{\{\hat{F}(t) - F(t)\}^2}{F(t)\{1-F(t)\}} \, dF(t) \qquad \text{A-D}$$

these can all be expressed as fn's of $U_{(i)}$ order st. of uniform see SM p328 & Fig 6.14 for an example

How to find $P_{H_0}\{T \ge t_{obs}\}$? – use tables of values, check R, simulate

- see also Exs 7.26, 7.27, 7.28 = SM

Example $Y_1, \dots, Y_n$ iid $N(\mu, \sigma_0^2)$ σ_0^2 known

$H_0: \mu = \mu_0$ $t(Y) = \dfrac{\sqrt{n}(\bar{y} - \mu_0)}{\sigma_0}$

under H_0 ~ $N(0,1)$ dist^n

$p_{obs} = pr(t(Y) \geq t_{obs}) = pr\left\{ N(0,1) \geq \dfrac{\sqrt{n}(\bar{y}_{obs} - \mu_0)}{\sigma_0} \right\}$

If alternative is $\mu \neq \mu_0$, then we would want to use T^2 or $|T|$, i.e. small & large value of T are evidence against H_0

Define 2-sided ~~test stat~~ $\cancel{p}$ pvalue = $\cancel{p}$ min (p^+_{obs}, p^-_{obs})

where $p^+_{obs} = Pr_{H_0}(T \geq t_{obs})$ & $p^-_{obs} = Pr_{H_0}(T \leq t_{obs})$

If T has a cont^s dist^n ~~then p = 2~~ sign level = $Pr(P \leq p_{obs}; H_0)$

~~$Pr(R \leq p^\circ) = 2p^\circ$~~

Example 7.25 & 7.26

7.25 t-test sig. level = 2 x (min)

7.26 $Y_1, \dots, Y_n$ iid F $H_0: F(\frac{1}{2}) = \mu_0$

$$T = \sum_{i=1}^{n} 1(Y_i > \mu_0) \sim Bin(n, \tfrac{1}{2}) \text{ under } H_0$$

$$p^+_{obs} = \sum_{r=t_{obs}}^{n} \binom{n}{r}\left(\frac{1}{2}\right)^n \qquad p^-_{obs} = \sum_{r=0}^{t_{obs}} \binom{n}{r}\left(\frac{1}{2}\right)^n$$

2-sided pvalue = ∴ need an example

$\simeq 2\min(p^+, p^-)$ bec. dist^n of T is symmetric

This is a special case ($k=2$) of the following:

Suppose we conduct k tests of sig, we p-values

$p_{obs,1}, \ldots, p_{obs,k}$, but report $p'_{obs} = \min(\ ,\ \ldots,\ )$

Then $\mathrm{pr}(P'_{obs} \leq p'_{obs}) = 1 - \mathrm{pr}(P'_{obs} > p'_{obs})$

$$\underset{\text{indep.}}{=} 1 - (1 - p'_{obs})^k \simeq k p'_{obs}$$

see CH p78 for an argument that is an upper bd. if tests are dependent.

Example $1-(1-0.05)^5 = 0.226$