1 Oct 01

Example. Two normals

$$p: N(\mu_0, \sigma_0^2) \qquad g: N(\mu, \sigma^2)$$

$$\theta = (\mu, \sigma)$$

$$K(\theta_0; \theta) =$$

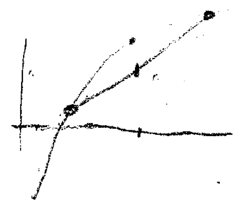
$$\frac{1}{2}\log\frac{\sigma_0^2}{\sigma^2} + \frac{1}{2}\left[\frac{\sigma^2}{\sigma_0^2} - 1 + \frac{(\mu_0 - \mu)^2}{\sigma^2}\right]$$

Note: 0 iff equality

Proof. (6)

$E \log X \le \log EX$

log concave



1 Oct 01

$\frac{\log a + \log b}{2} \le \log\left(\frac{\alpha+\beta}{2}\right)$

Geo mean vs Ari

$$E_p \log \frac{g}{p} \le \log E_p \frac{g}{p} = 0$$

Jensen

For equality $p = g$ ae

$$\int p \log \frac{g}{p}\, dy \le 0$$

$$\int p \log \frac{p}{g}\, dy \ge 0$$

So the estimate $\hat{\theta}$ is looking for the member of the family $\{f_i(y|\theta)\}$ closest to P measuring distance by discrimination information.

If P is in the family, i.e. $f_i(y|\theta_*)$ then

$$\int f_i(y|\theta_*) \log f_i(y|\theta)\, dy$$

maximized at $\theta = \theta_*$

$$\int f_i(y|\theta_*) \log f_i(y|\theta)\, dy \le \int f_i(y_i|\theta_*) \log f_i(y|\theta_*)\, dy$$

1 Oct 01

Example. n i.i.d.

Suppose: $f(y|\theta) = \dfrac{\theta^y e^{-\theta}}{y!}$ $y = 0, 1, \ldots,$

Poisson

True P: $\binom{N}{y} \pi^y (1-\pi)^{N-y}$

Binomial

$$-\log f(y|\theta) \sim -(y \log \theta - \theta)$$

$$\psi(y|\theta) = -\left(\frac{y}{\theta} - 1\right)$$

$$E_p \psi(Y|\theta) = -\left(\frac{N\pi}{\theta} - 1\right)$$

Equals 0 at $\theta_* = N\pi$

So $\hat{\theta} = \bar{y}$

var $\hat{\theta}$ $N\pi($

$\hat{\theta} \xrightarrow{P} N\pi$

10 oct 02

$$\mathrm{var}\ \hat{\theta} = \frac{N\pi(1-\pi)}{n}$$

would have thought $\frac{\theta}{n}$

OK for π small

Will return to K-L when developing AIC etc.

3 Oct, 2001

Maximum Likelihood Estimation

$f(y|\theta)$ p.m.f. or p.d.f. for a r.v. Y
θ unknown parameter

Data: $y_1, \dots, y_n$

Likelihood function

$L(\theta) = f(y_1|\theta) \dots f(y_n|\theta)$ independent

It is a stochastic process.

MLE

$\hat{\theta}$ provides $\max_{\theta} L(\theta)$

same as $\max_{\theta} \log L(\theta)$

Under conditions it is: asymptotically consistent, normal and efficient.

3 Oct, 2001

Examples

1. **Poisson**. Consider the birth data for a specific day of the week

$$f(y|\theta) = \text{Prob}\{Y=y\}$$

$$= \frac{1}{y!}\theta^y e^{-\theta}, \quad y=0,1,2,\ldots$$

$$EY = \theta$$

$$L(\theta) = \prod_i \frac{1}{y_i!}\theta^{y_i} e^{-\theta}$$

$$\log L(\theta) = \sum_i \left(y_i \log\theta - \theta - \log y_i!\right)$$

$$\frac{\partial}{\partial\theta}: \quad \left(\sum_i y_i\right)\frac{1}{\theta} - n \qquad \text{Note } E = 0$$

$$\hat{\theta} = \bar{y}$$

$$\frac{\partial^2}{\partial\theta^2}: \quad -\left(\sum_i y_i\right)\frac{1}{\theta^2} \qquad \text{maximum}$$

$$nI(\theta) = -nE\left(\frac{\partial^2 \log L}{\partial\theta^2}\right) = n/\theta, \quad \text{var}\,\hat{\theta} \sim \theta/n$$

```
SUNDAY

N = 103   Median = 15
Quartiles = 12, 18

Decimal point is at the colon

    4 : 0
    5 :
    6 :
    7 : 0
    8 : 00
    9 : 0000
   10 : 0000000
   11 : 000000000
   12 : 000000000
   13 : 0000000000
   14 : 00000
   15 : 0000000000
   16 : 0000000000
   17 : 00000
   18 : 000000000
   19 : 00000000
   20 : 00000
   21 : 0
   22 : 000
   23 :
   24 : 0
   25 : 00
   26 : 0

MONDAY

N = 103   Median = 18
Quartiles = 15, 20

Decimal point is at the colon

    6 : 0
    7 :
    8 : 0
    9 :
   10 : 000
   11 : 0000
   12 : 00000
   13 : 00000
   14 : 00000
   15 : 0000000
   16 : 00000000
   17 : 00000000000
   18 : 00000000000
   19 : 00000000000
   20 : 0000000000
   21 : 000000
   22 : 0
   23 : 000
   24 : 0
   25 : 00000
   26 : 00
   27 : 0
   28 :
   29 : 0

High:  31
```

3 Oct 2001

2. Normal regression.

$$f(y \mid \alpha, \beta, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{(y-\alpha-\beta x)^2}{2\sigma^2} \right\} \qquad -\infty < y < \infty$$

$$\log L(\alpha,\beta,\sigma) = \sum_i \left(-\log\sigma - \frac{1}{2\sigma^2}(y_i - \alpha - \beta x_i)^2 - \log\sqrt{2\pi} \right)$$

$$= -n\log\sigma - \frac{1}{2\sigma^2}\sum_i (y_i - \alpha - \beta x_i)^2 - n\log\sqrt{2\pi}$$

$\frac{\partial}{\partial\alpha}, \frac{\partial}{\partial\beta}$: OLS!

$$\frac{\partial}{\partial\sigma} : \quad -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_i (y_i - \alpha - \beta x_i)^2$$

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

Reparametrize $\quad y_i - \alpha^* - \beta(x_i - \bar{x})$

$$\hat{\alpha}^* = \bar{y}, \quad \hat{\beta} = \sum_i (x_i - \bar{x}) y_i \Big/ \sum_i (x_i - \bar{x})^2$$

3 Oct 2001

$$\frac{\partial}{\partial \alpha^*}: \quad \frac{1}{\sigma^2} \sum_i \left(y_i - \alpha^* - \beta(x_i - \bar{x}) \right)$$

$$= \frac{1}{\sigma^2} n (\bar{y} - \alpha^*)$$

$$\frac{\partial}{\partial \beta}: \quad \frac{1}{\sigma^2} \sum_i \left(y_i - \alpha^* - \beta(x_i - \bar{x}) \right)(x_i - \bar{x})$$

$$= \frac{1}{\sigma^2} \left(\sum_i (x_i - \bar{x}) y_i - \sum_i (x_i - \bar{x})^2 \beta \right)$$

$$\frac{\partial^2}{\partial \alpha^{*2}}: \quad -\frac{n}{\sigma^2}$$

$$\frac{\partial^2}{\partial \alpha^* \partial \beta}: \quad 0$$

$$\frac{\partial^2}{\partial \alpha^* \partial \sigma}: \quad -\frac{2}{\sigma^3} n (\bar{y} - \alpha^*) \qquad E_0 = 0$$

$$\frac{\partial^2}{\partial \beta^2}: \quad -\frac{1}{\sigma^2} \sum_i (x_i - \bar{x})^2$$

$$\frac{\partial^2}{\partial \beta \partial \sigma}: \quad -\frac{2}{\sigma^2} \left(\sum_i (x_i - \bar{x}) y_i - \sum_i (x_i - \bar{x})^2 \beta \right) \qquad E_0 = 0$$

$$\frac{\partial^2}{\partial \sigma^2}: \quad \frac{n}{\sigma^2} - \frac{3}{\sigma^4} \sum_i (y_i - \alpha - \beta x_i)^2 \qquad E_0 = -\frac{2n}{\sigma^2}$$

3 Oct. 2001

$$I(\alpha, \beta, \sigma)_0 = \begin{bmatrix} \frac{n}{\sigma^2} & 0 & 0 \\ 0 & \frac{\sum (x_i - \bar{x})^2}{\sigma^2} & 0 \\ 0 & 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

The limiting conditions.

$$\gamma_i(\alpha^*, \beta, \sigma) = E\, \rho_i(Y_i \mid \alpha^*, \beta, \sigma)$$

$$\underset{\sim}{\lambda}_i(\theta) = E\, \underset{\sim}{\psi}_i(Y_i \mid \theta)$$

$$\begin{bmatrix} -\frac{n}{\sigma^2}(\alpha_0^* - \alpha^*) \\ -\frac{(\beta_0 - \beta)\sum (x_i - \bar{x})^2}{\sigma^2} \\ -\frac{n}{\sigma^3}(\alpha_0^* - \alpha^*)^2 - \frac{(\beta_0 - \beta)^2}{\sigma^3}\sum (x_i - \bar{x})^2 \end{bmatrix}$$

Want $\frac{1}{n}\sum_i (x_i - \bar{x})^2 \longrightarrow \sigma_x^2$

3 Oct. 01

3. Example from seismology

Fault plane estimation from signs of first motion

Fault plane is described by angles Θ_T, Φ_T, Θ_P

$i = 1, \ldots, I$ indexes events

$j = 1, \ldots, J_i$ indexes stations within events

$A_{ij}(\Theta_T, \Phi_T, \Theta_P)$: theoretical signed amplitude

Data: $Y_{ij} = \begin{matrix} 1 \\ -1 \end{matrix}$

Signal: $s(t)$

Seismogram: $Z(t) = s(t) + \epsilon(t)$

"Y" $= +1 \qquad Z(t_0) > 0$

$\qquad -1 \qquad Z(t_0) < 0$

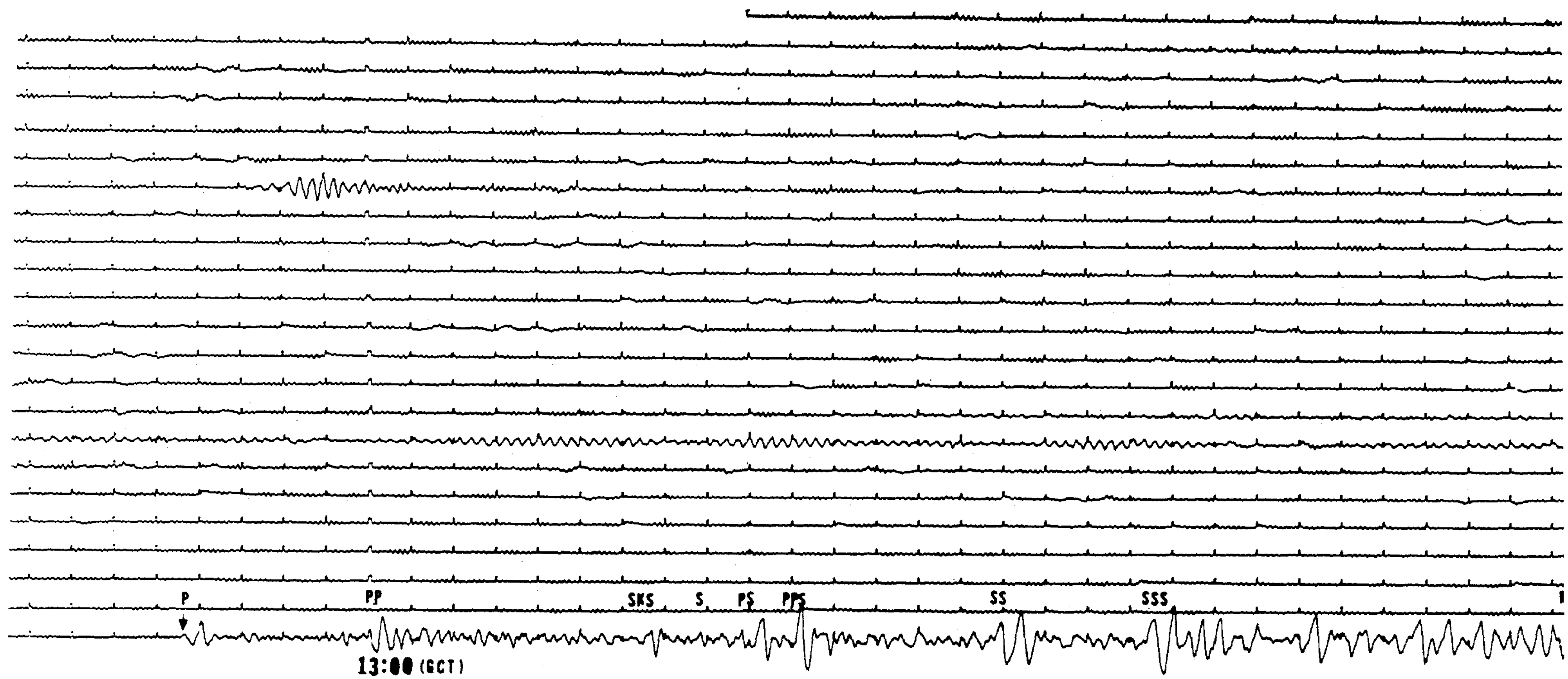


Figure 3.4 An elegant earthquake record (bottom line of the seismogram) made at the Berkeley Observatory on a standard modern seismograph. The record gives the vertical motion of the ground surface. The interval between the tick marks on the record corresponds to 1 minute, and time increases from left to right. This earthquake occurred near Borneo at a distance of 11,000 km from Berkeley. The onset of *P* waves is clearly seen, together with the *PP* reflection. This is followed by the onset of *SKS* and *S* waves and the reflections *PS*, *PPS*, *SS*, and *SSS*. At the end of the bottom trace can be seen the Rayleigh wave train, starting with long but decreasing periods (an example of wave dispersion). The record is not complete because the wave motion was interrupted by the operator inadvertently changing the seismogram.

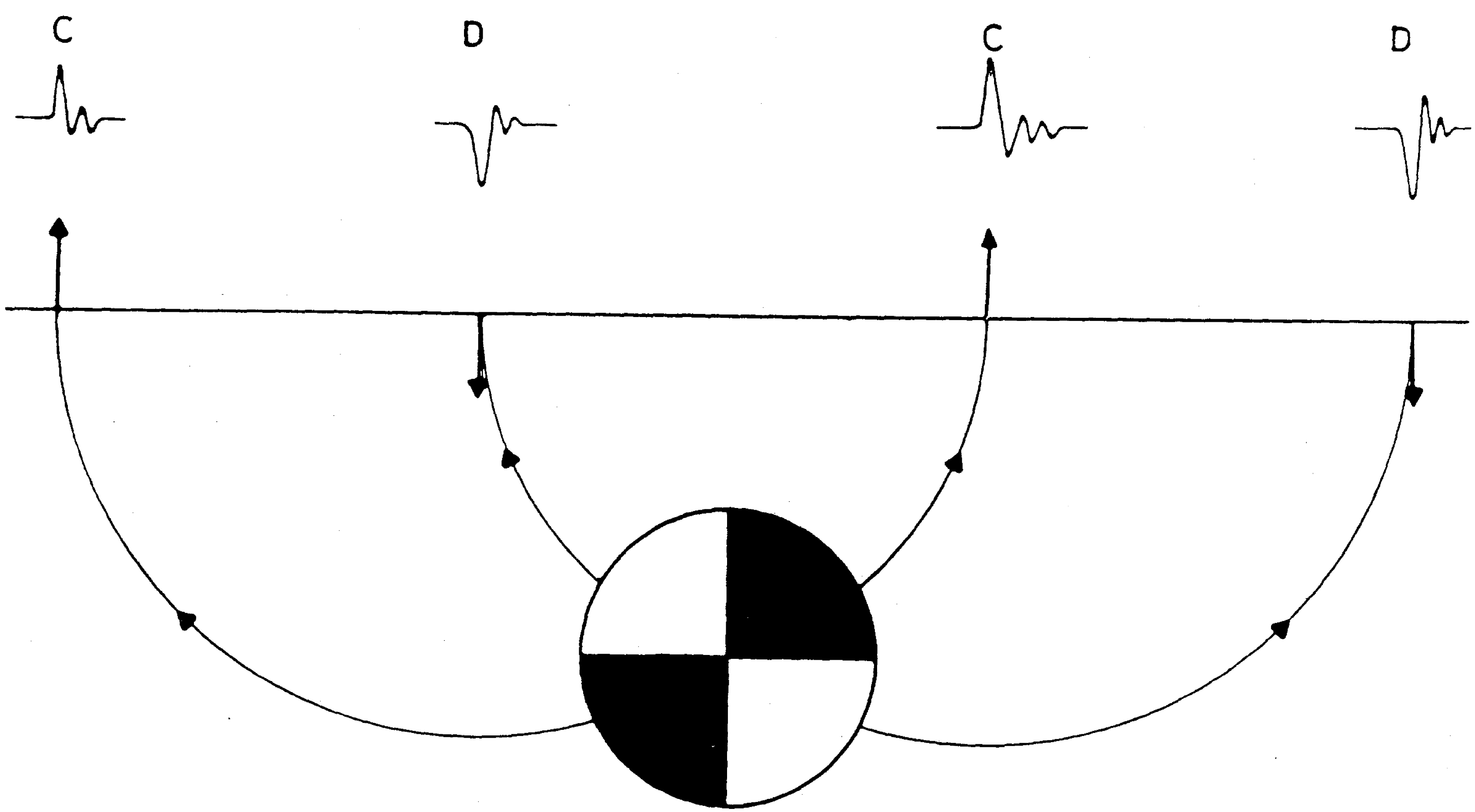
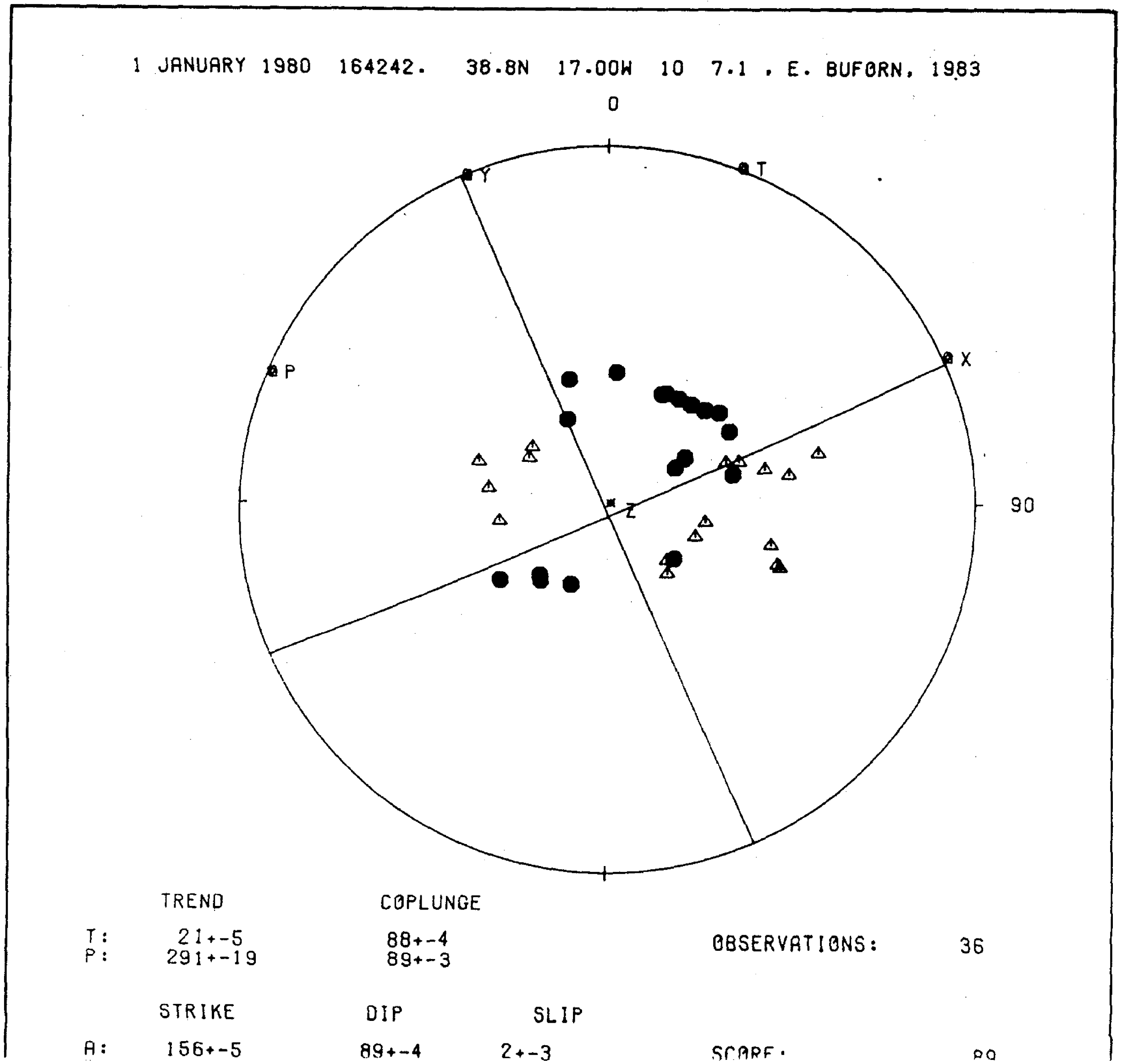


Figura 7.—Distribución de compresiones (cuadrantes en negro) y dilataciones (en blanco) sobre la esfera focal y desplazamientos de la onda *P* en la superficie de la Tierra.


1 JANUARY 1980 164242. 38.8N 17.00W 10 7.1 , E. BUFØRN, 1983
0
90
P
Y
T
X
Z
TREND
CØPLUNGE
T: 21+-5 88+-4
P: 291+-19 89+-3
ØBSERVATIØNS: 36
STRIKE
DIP
SLIP
A: 156+-5 89+-4 2+-3
SCØRE: 89

regional stress pattern.

As starting values for our program, we adopted for the three independent angles (θ_T, Φ_T, Φ_P = 45°, 315°, 135°) and $\rho_i = 5$ for all events. The initial value of the logarithm of the likelihood function, given by (11), is $-L = 1322$. After 30 iterations the program yielded 39°.0, 318°.3, 138°.9 for the axes angles and $-L = 407$. The final solution is

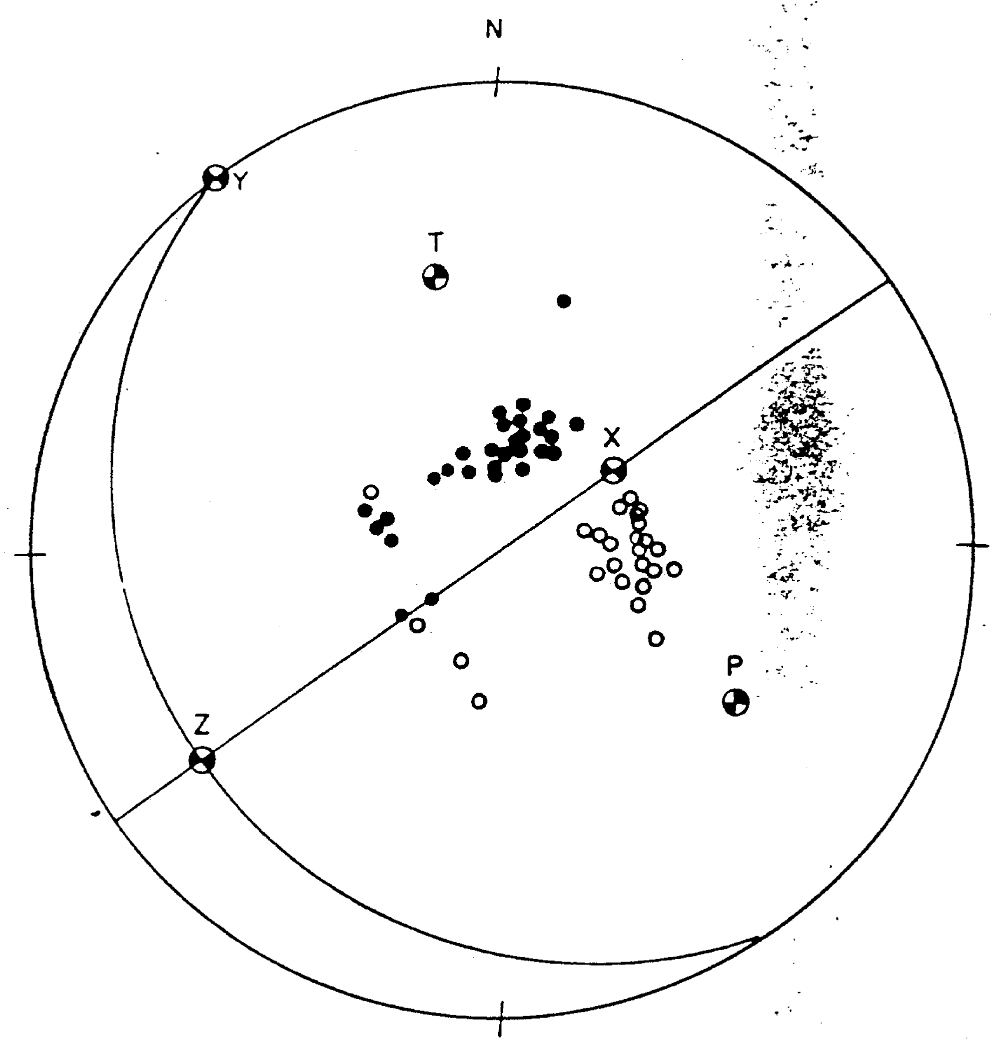


FIG. 4. *P* first-motion data for earthquake number 4 of the Alaska sequence, March 30, 1964, 2h.

3 Oct. 01

Probability of reader/recorder error: γ

$$\text{Prob}(Y=1) = \gamma + (1-2\gamma)\,\text{Prob}\{Z(t_0)>0\}$$

$$\text{Prob}\{Z(t_0)>0\} = \text{Prob}\{\epsilon(t_0) > -\alpha A\}$$

$$= \Phi(\rho A) \qquad \rho = \alpha/\sigma$$

where $s(t_0) = \alpha A$

$$\pi_{ij} = \text{Prob}\{Y_{ij}=1\}$$

Likelihood

$$L(\Theta_T, \Phi_T, \Theta_P, \rho, \gamma)$$

$$= \prod_{i,j} \pi_{ij}^{(1+y_{ij})/2} (1-\pi_{ij})^{(1-y_{ij})/2}$$

4 Oct. 2001

For the mle

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

What about efficiency?

Cramér-Rao Inequality

~~Statistic~~ $\overset{k\times 1}{\underset{\sim}{T}} = \underset{\sim}{T}(\underset{\sim}{y})$ $\qquad \underset{\sim}{T}_n$

Bias $\overset{k\times 1}{\underset{\sim}{b}(\theta)} = E\overset{k\times 1}{\underset{\sim}{T}} - \overset{k\times 1}{\underset{\sim}{\theta}}$ $\qquad : \underset{\sim}{b}_n(\theta)$

$$E_\theta\{(\underset{\sim}{T} - \underset{\sim}{\theta})(\underset{\sim}{T} - \underset{\sim}{\theta})^T\} \geq \frac{1}{n}\left(\underset{\sim}{I} + \frac{\partial \underset{\sim}{b}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right) \bar{\underset{\sim}{I}}(\theta)^{-1}\left(\underset{\sim}{I} + \frac{\partial \underset{\sim}{b}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right)^T$$

$$+ \underset{\sim}{b}(\underset{\sim}{\theta})\underset{\sim}{b}(\underset{\sim}{\theta})^T$$

$\bar{\underset{\sim}{I}}(\theta) = \sum_i \mathcal{I}_i / n$

If $\underset{\sim}{T}$ is unbiased

$$E_\theta\{(\underset{\sim}{T} - \underset{\sim}{\theta})(\underset{\sim}{T} - \underset{\sim}{\theta})^T\} \geq \frac{1}{n}\bar{\underset{\sim}{I}}(\underset{\sim}{\theta})^{-1}$$

4 Oct 2001

Proof.

$$\int (\underset{\sim}{T} - \underset{\sim}{\theta} - \underset{\sim}{b}_n(\theta)) f_n(\underset{\sim}{y}|\theta) d\underset{\sim}{y} = \underset{\sim}{0}$$

$k \times 1$ $\qquad$ $k \times 1$

$\frac{\partial}{\partial \underset{\sim}{\theta}}$:

$$\int \left(-\underset{\sim}{I} - \frac{\partial}{\partial \theta} b_n(\underset{\sim}{\theta})\right) f_n(\underset{\sim}{y}|\underset{\sim}{\theta}) d\underset{\sim}{y} + \int (\underset{\sim}{T} - \underset{\sim}{\theta} - \underset{\sim}{b}_n(\underset{\sim}{\theta})) \frac{\partial \log f_n(\underset{\sim}{y}|\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}} f_n(\underset{\sim}{y}|\underset{\sim}{\theta}) d\underset{\sim}{y} = \underset{\sim}{0}$$

$k \times k$ $\qquad$ $k \times 1$ $\qquad$ $1 \times k$ $\qquad$ $k \times k$

$$\int (\underset{\sim}{T} - \underset{\sim}{\theta} - \underset{\sim}{b}_n(\underset{\sim}{\theta})) \frac{\partial \log f_n(\underset{\sim}{y}|\underset{\sim}{\theta})}{\partial \theta} f_n(\underset{\sim}{y}|\underset{\sim}{\theta}) d\underset{\sim}{y}$$

$$= E\left\{ \underbrace{(\underset{\sim}{T} - \underset{\sim}{\theta} - \underset{\sim}{b}_n(\theta))}_{Y} \underbrace{\frac{\partial \log f_n(\underset{\sim}{y}|\theta)}{\partial \underset{\sim}{\theta}}}_{X} \right\}$$

$$= \underset{\sim}{I} + \frac{\partial}{\partial \theta} \underset{\sim}{b}_n(\underset{\sim}{\theta}) \qquad (*)$$

$$EX^2 EY^2 \geq |EXY|^2$$

Vector case

$$E \underset{\sim}{Y}\underset{\sim}{Y}^T \geq E \underset{\sim}{Y}\underset{\sim}{X}^T (E\underset{\sim}{X}\underset{\sim}{X}^T)^{-1} E\underset{\sim}{X}\underset{\sim}{Y}^T$$

(from conditional distn of normal) works generalized inverse too

4 Oct 2001

So, from ⊛,

$$\operatorname{var} \underset{\sim}{T} \geq \left(\underset{\sim}{I} + \frac{\partial}{\partial \theta} \underset{\sim}{b}_n(\underset{\sim}{\theta})\right)\left(\operatorname{var} \frac{\partial \log f_n}{\partial \theta}\right)^{-1}\left(\underset{\sim}{I} + \frac{\partial}{\partial \theta} \underset{\sim}{b}_n(\underset{\sim}{\theta})\right)^T$$

$$\operatorname{var} \frac{\partial \log f_n}{\partial \theta} = \sum_i \underset{\sim}{I}_i(\underset{\sim}{\theta}) = n \bar{\underset{\sim}{I}}(\underset{\sim}{\theta})$$

Final step: $E\{(\underset{\sim}{T}-\underset{\sim}{\theta})(\underset{\sim}{T}-\underset{\sim}{\theta})^T\} = \operatorname{var} \underset{\sim}{T} + \underset{\sim}{b}(\underset{\sim}{\theta})\, \underset{\sim}{b}(\underset{\sim}{\theta})^T$

Comments.

When $n \to \infty$, see mle asymptotically efficient in a sense. [There are various senses, eg Hajek]

5 Mar 04

Review

Nonlinear estimation

M-estimates (Huber, ...)

e.g. nonlinear regression
maximum likelihood
robust/resistant
. . .

Stochastic model, parameter θ, $\dim(\theta) < \infty$

Data y_i

Loss function $\rho_i(y|\theta)$

Estimate, $\hat{\theta}$

$$\min_{\theta} \sum_{i=1}^{n} \rho_i(y_i|\theta)$$

OR

$$\sum_{i=1}^{n} \underset{\sim}{\psi}_i(y_i|\hat{\theta}) = \underset{\sim}{0} \qquad \text{estimating eq}^n$$

"True" value θ_0

Theorem

$$\sqrt{n}\,(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \underset{\sim}{\Lambda}^{-1} \underset{\sim}{C} \underset{\sim}{\Lambda}^{-1}) \qquad \text{as } n \to \infty$$

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \qquad \text{consistent}$$

5 March 04

Conditions / definitions.

$$\gamma_i(\theta) = E\{\rho_i(Y_i|\theta)\}$$

$$\psi_i(y_i|\theta) = \frac{\partial}{\partial\theta}\rho_i(y_i|\theta)$$

$$\underset{\sim}{\lambda}_i(\theta) = E\{\underset{\sim}{\psi}_i(Y_i|\theta)\}$$

$$\frac{1}{n}\sum_i \gamma_i(\theta) \rightarrow \gamma(\theta)$$

$$\frac{1}{n}\sum_i \underset{\sim}{\lambda}_i(\theta) \rightarrow \underset{\sim}{\lambda}(\theta)$$

$$\gamma(\theta) > \gamma(\theta_0) \quad \forall\, \theta \neq \theta_0$$

$$\underset{\sim}{\lambda}(\theta_0) = \underset{\sim}{0}, \quad |\underset{\sim}{\lambda}(\theta)| > 0 \quad \forall\, \theta \neq \theta_0$$

$$\underset{\sim}{C}_i = \mathrm{var}\{\underset{\sim}{\psi}_i(Y_i|\theta_0)\}$$

$$\underset{\sim}{\Lambda}_i = \left.\frac{\partial \underset{\sim}{\lambda}_i(\theta)}{\partial\theta}\right|_{\theta_0}$$

$$\frac{1}{n}\sum_i \underset{\sim}{C}_i \rightarrow \underset{\sim}{C}$$

$$\frac{1}{n}\sum_i \underset{\sim}{\Lambda}_i \rightarrow \underset{\sim}{\Lambda}$$

5 March 04

Nonlinear regression.
Assumed model

$$Y_i = \mu_i(\theta) + \epsilon_i$$

$$\epsilon_i \sim \text{indep, mean } 0, \text{ variance } \sigma^2$$

$$\rho_i(y|\theta) = [y - \mu_i(\theta)]^2 \qquad \text{LS}$$

$$\underset{\sim}{\psi}_i(y|\theta) = -2[y - \mu_i(\theta)] \frac{\partial \mu_i(\theta)}{\partial \theta}$$

Suppose:
data (y_i, x_i)

$$\mu_i(\theta) = \mu(x_i, \theta)$$

$$E\{Y_i | x_i\} = g(x_i)$$

$$\frac{1}{n} \sum_i \rho_i(y_i|\theta) \xrightarrow{p} \gamma(\theta) = \sigma^2 + \int [g(x) - \mu(x,\theta)]^2 dF_X(x)$$

$$\theta_0 = \arg\min_\theta \int [g(x) - \mu(x,\theta)]^2 dF_X(x)$$

Model true, $g(x) = \mu(x, \theta_0)$

5 March 04

$$\frac{1}{n}\sum_i \underset{\sim}{\psi}_i(y_i|\theta) \xrightarrow{p} \underset{\sim}{\lambda}(\theta)$$

$$= -2\int [g(x)-\mu(x,\theta)]\frac{\partial\mu(x,\theta)}{\partial\theta}\, dF_X(x)$$

$$\int [g(x)-\mu(x,\theta_0)]\left.\frac{\partial\mu(x,\theta)}{\partial\theta}\right|_{\theta_0} dF_X(x) = \underset{\sim}{0}$$

$$\underset{\sim}{C}_i = \mathrm{var}\{\underset{\sim}{\psi}_i(Y_i|\theta_0)\}$$

$$= 4\sigma^2 \frac{\partial\mu(x_i,\theta_0)}{\partial\theta_0}\frac{\partial\mu(x_i,\theta_0)}{\partial\theta_0}^T$$

$$\frac{1}{n}\sum_i \underset{\sim}{C}_i \to \underset{\sim}{C} = 4\sigma^2\int \frac{\partial\mu(x,\theta_0)}{\partial\theta_0}\frac{\partial\mu(x,\theta_0)}{\partial\theta_0}\, dF_X(x)$$

$$\frac{1}{n}\sum_i \underset{\sim}{A}_i \to \underset{\sim}{A} = \frac{\partial\underset{\sim}{\lambda}(\theta_0)}{\partial\theta_0}$$

$$= 2\int \frac{\partial\mu(x,\theta_0)}{\partial\theta_0}\frac{\partial\mu(x,\theta_0)}{\partial\theta_0}^T dF_X(x)$$

$$-2\int [g(x)-\mu(x,\theta_0)]\frac{\partial^2\mu(x,\theta_0)}{\partial\theta_0\partial\theta_0^T}\, dF_X(x)$$

$$= \frac{\underset{\sim}{C}}{2\sigma^2} \quad \text{if } g(x) = \mu(x,\theta_0)$$

5 March 04

Maximum likelihood.

Assumed model

Y_i has pdf (or pmf) $f_i(y|\theta)$

$$\rho(y|\theta) = -\log f_i(y|\theta)$$

$$\underset{\sim}{\psi}_i(y|\theta) = -\frac{\partial}{\partial\theta}\log f_i(y|\theta)$$

$$= -\frac{1}{f_i(y|\theta)}\frac{\partial f_i(y|\theta)}{\partial\theta}$$

True distribution P

$$\frac{1}{n}\sum_i \rho_i(y_i|\theta) \xrightarrow{p} -\lim \frac{1}{n}\sum_i E_P\{\log f_i(Y|\theta)\}$$

$$= \gamma(\theta)$$

$$\theta_0 = \arg\max_\theta \int \log f_i(y|\theta)\,dP(y)$$

$$\frac{1}{n}\sum_i \underset{\sim}{\psi}_i(y_i|\theta) \xrightarrow{p} -\lim \frac{1}{n}\sum_i E_P\left\{\frac{\partial \log f_i(Y|\theta)}{\partial\theta}\right\}$$

$$= \underset{\sim}{\lambda}(\theta)$$

θ_0: $\quad E_P\left\{\frac{\partial \log f_i(Y|\theta_0)}{\partial\theta}\right\} = \underset{\sim}{0}$

5 March 04

$$\underset{\sim}{C}_i = \mathrm{var}_p\, \underset{\sim}{\psi}_i(Y_i|\theta)$$

$$= E_p\left\{ \frac{\partial \log f_i(Y|\theta)}{\partial \theta} \frac{\partial \log f_i(Y|\theta)}{\partial \theta}^T \right\}$$

$$\underset{\sim}{C} = \lim \frac{1}{n} \sum_i E_p\left\{ \frac{\partial \log f_i(Y|\theta)}{\partial \theta} \frac{\partial \log f_i(Y|\theta)}{\partial \theta}^T \right\}$$

$$\lambda_i(\theta) = - E_p\left\{ \frac{\partial^2 \log f_i(Y|\theta)}{\partial \theta\, \partial \theta^T} \right\}$$

$$\underset{\sim}{\Lambda} = -\lim \frac{1}{n} \sum_i E_p\left\{ \frac{\partial^2 \log f_i(Y|\theta)}{\partial \theta\, \partial \theta^T} \right\}$$

If $\quad p(y) = f_i(y|\theta_0)$

$$E_p\left\{ \frac{\partial \log f_i(y|\theta)}{\partial \theta} \right\} = \int \frac{\partial \log f_i(y|\theta)}{\partial \theta} f_i(y|\theta_0)\, dy$$

$$= 0 \quad \text{at } \theta = \theta_0$$

5 March 04

K-L "distance"

$$K(p;q) = \int p(y) \log \frac{p(y)}{q(y)}\, dy$$

$$K(p;q) > 0$$
$$= 0 \quad \text{iff} \quad p \equiv q \quad \text{a.e.}$$

A measure of closeness of q to p.

Consider

$$K(p;f_i) = \int p(y) \log \frac{p(y)}{f_i(y|\theta)}\, dy$$

$$= \int p(y) \log p(y)\, dy - \int p(y) \log f_i(y|\theta)\, dy$$

$$\theta_0 = \arg\max_\theta \int p(y) \log f_i(y|\theta)\, dy$$

$$\int f_i(y|\theta_0) \log f_i(y|\theta)\, dy$$

maximized at $\theta = \theta_0$

5 Mar 04

Conclusion

Estimate, $\hat{\theta}$, useful if model "close enough" in one of these "distances".