2060-26

Advanced School on Non-linear Dynamics and Earthquake Prediction

28 September - 10 October, 2009

Space-Time Earthquake prediction:
The Error Diagrams

Molchan G.M.

IIEPT
Moscow
Russia

Advanced School on the Non-linear Dynamics
and Earthquake Prediction
28 September - 10 October 2009
The Abdus Salam International Centre for Theoretical Physics, Trieste

Space-Time Earthquake prediction:

The Error Diagrams

MOLCHAN G.M.

IIEPT, Moscow, Russia

Papers: GJI, 173 (2008), 1012-1017
PAGEOPH (2009), accepted

OUTLINE

- The simplest space-time prediction problem

  Generalized error diagrams

- Statistical problems
- Prediction efficency at the research stage

MOTIVATION

The main prediction characteristics:

- n rate of failures-to-predict
- τ rate of alarm time

Nonuniqueness of τ for space-time prediction:

$$\tau = \frac{1}{T}\int_0^T \mu(G_t)dt, \quad T >> 1$$

G_t – alarm zone at t

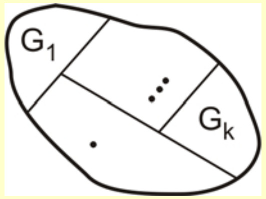
$$\mu(G) = \begin{cases} \text{area of } G \\ \text{rate of target events in } G \\ \text{population in } G \end{cases}$$

$$\Downarrow$$

Non-adequate interpretation of predictability in terms of (n, τ)

THE SIMPLEST SPACE-TIME PREDICTION PROBLEM

NOTATION



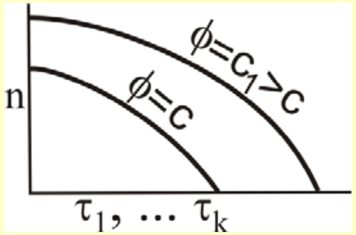
$G = \bigcup_{1}^{k} G_i$ – area, *I*(*t*) – data for prediction at *t*

$\pi = \{\pi(t)\}$ – **Strategy** based on *I*(*t*)

Solution $\pi(t \,|\, G_i) = \begin{cases} \text{alarm} \\ \text{no alarm} \end{cases}$ in $G_i \times (t, t+\Delta), \quad \Delta << 1$

$(n, \tau_1, \ldots \tau_k)$ **– Errors**, τ_i – rate of alarm time in G_i

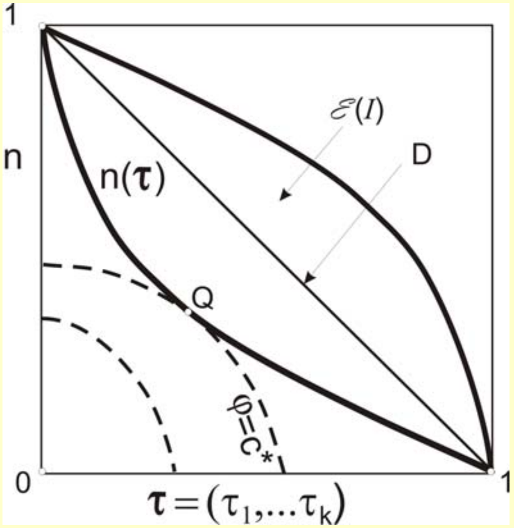
Goal: $\varphi(n, \tau_1, \ldots, \tau_k) \Rightarrow min$



PROBLEM: π^* **(**optimal strategy**) – ?**

$\inf_{\pi} \varphi$ **– ?**

SOLUTION: the simplest prediction problem



$\mathcal{E}(I) = \{(n, \tau_1 ... \tau_\kappa)_\pi\}$ error set, convex

$D = \{n+\tau_\lambda=1\}$ trivial strategies (random guessing)

$$\tau_\lambda = \sum_{i=1}^{n} \tau_i \lambda_i / \Lambda$$

λ_i	rate of target events in	G_i
$\Lambda = \sum \lambda_i$		G

$n=n(\tau_1 \dots \tau_k)$ – correct version of the **error diagram**

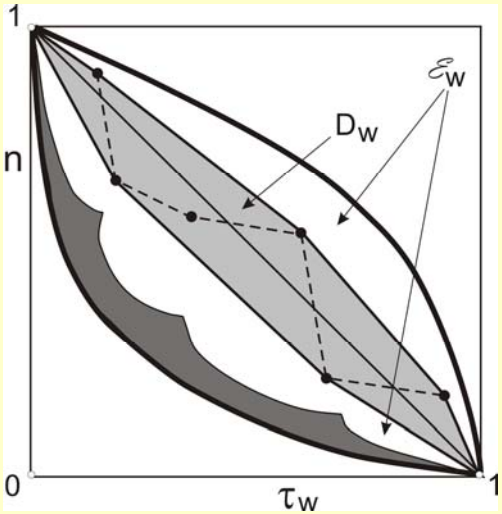
- $\inf_{\pi} \varphi = \varphi(Q)$
- $\boxed{\pi^*(t|G_i) = \text{alarm}}$ in $G_i \times (t, t+\Delta)$ as soon as $\boxed{r(t|G_i) \geq r_i^0}$

$r(t|G_i) = P\{dN(t|G_i) = 1 | I(t)\} / \Delta$ hazard function for G_i

$$\boxed{r_i^0 = b_i / a \quad \text{for} \quad \varphi = a\Lambda n + \Sigma \tau_i b_i}$$

- $\boxed{n(\tau, ... \tau_k) = \Sigma \lambda_i n(\tau_i | G_i) / \Lambda}$

REDUCED ERROR DIAGRAM



$$\tau_w = \sum_{i=1}^{k} w_i \tau_i, \quad \sum w_i = 1$$

$$\gamma : (n, \tau_1 ... \tau_k) \to (n, \tau_w)$$

$\mathcal{E} \to \mathcal{E}_w$

$D \to D_w$ **(**grey zone**)**

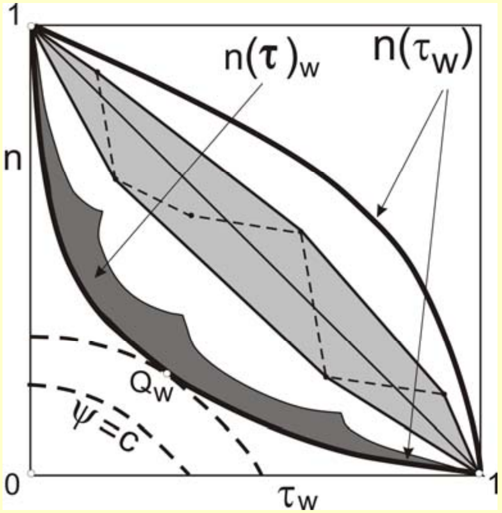
$n(\tau) \to n(\tau)_w$ **(**dark zone**)**

- D_w – convex hull of the points (n, τ_w):
  $n = 1 - \sum \lambda_i \varepsilon_i / \Lambda, \; \tau_w = \sum w_i \varepsilon_i, \; \varepsilon_i = \{0,1\}$
- $w_1 = ... = w_k \Rightarrow$

lower boundary of D_w = convex minorant of the points:

(n,τ_w): (1, 0), $(1 - \lambda_{(k)} / \Lambda), 1/k), ..., (1 - \sum_{i=1}^{p} \lambda_{(k-i+1)} / \Lambda, p/k)...(0,1)$

$\lambda_{(1)} \le \lambda_{(2)} ... \le \lambda_{(k)}$,ordered $\{\lambda_i\}$



- D_w, grey zone ={$n+\tau_w=1$} ↔ $w_i=\lambda_i/\Lambda$, $i=1,\dots,k$

- {$n_i(\tau)$} are picewise smooth, $i=1,\dots,k$:

$$n(\tau)_w\text{, dark zone} = n(\tau_w) \leftrightarrow I(t)\text{ is trivial}$$

- upper bound of $n(\tau)_w$:

$$n^+(x)=\max_{i,\varepsilon}\{\lambda_i/\Lambda\cdot n_i(x/w_i-a_i(\varepsilon))+b_i(\varepsilon)\}$$

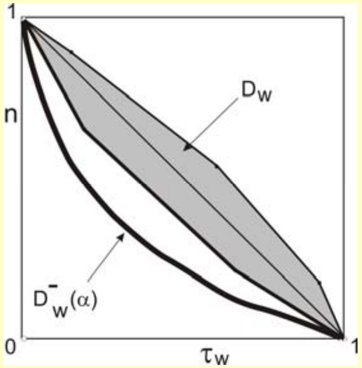
$$\varepsilon=(\varepsilon_1\dots\ \varepsilon_k),\ \ \varepsilon_i=(0,1)$$

$$a_i(\varepsilon)=\sum_{j\neq i}w_j\varepsilon_j/w_i,\qquad b_i(\varepsilon)=\sum_{j\neq i}\lambda_j(1-\varepsilon_j)/\Lambda$$

- optimal property of $n(\tau_w)$:

$$\varphi(n,\tau_1\dots\tau_k)=\psi(n,\tau_w)\Rightarrow\min_{\pi}\varphi=\psi(Q),\ \ Q\in n(\tau_w)$$

STATISTICAL PROBLEMS



1. Confidence zone for D_w of level α, $D_w(\alpha)$

PROBLEM: find $D_W(\alpha)$ **using a confidence zone** $U(p|\alpha)$

for unknown $\{p_i=\lambda_i/\Lambda\}$

SOLUTION

◆ $U(p|\alpha)$: $\sum\limits_{i=1}^{k}(p_i-\hat{p}_i)^2/\hat{p}_i < q(\alpha)$

$\{\hat{p}_i\}$, unbiased estimates of $\{p_i\}$, based on $N\_$ events, $\hat{p}_i N\_ >> 1$

$$q(\alpha)\approx[(k-1)+\rho\sqrt{2(k-1)}/N\_$$

$$\rho(\alpha)\approx 2.5 \text{ for } k\geq 5,\ \alpha>97.5\%$$

◆ $D_w^-(\alpha)$ is the convex minorant of (n, τ_w)

- $n(\varepsilon)=f(\sum\limits_{i=1}^{k}\hat{p}_i\varepsilon_i)$, $\quad\tau_w(\varepsilon)=\sum\limits_{i=1}^{k}w_i\varepsilon_i$, $\quad\varepsilon_i=\begin{cases}0\\1,\end{cases}$ $\#\{\varepsilon\}=2^k$

  $$f(x)=\begin{cases}1-x-\sqrt{qx(1-x)}, & (1+q)x<1\\ 0\qquad\qquad\quad , & (1+q)x>1\end{cases}$$

- $n(j)=f(\sum\limits_{i=1}^{j}\hat{p}_{(k-i+1)})$, $\tau_w(j)=j/k$, $j=1,\ldots k$ if $w_1=\ldots=w_k$

  $\hat{p}_{(1)}\leq\hat{p}_{(2)}\leq\ldots\leq\hat{p}_{(k)}$, ordered $\{\hat{p}_i\}$

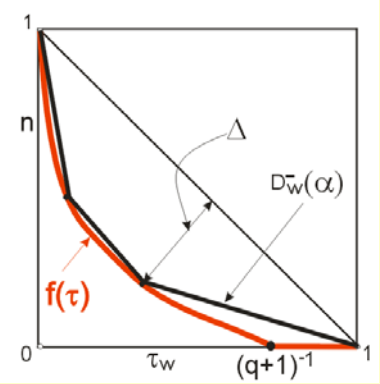
STATISTICAL PROBLEMS (continuation)

♦ Confidence α-zone for $\{n+\tau_\lambda=1\}$

$$w_i=\hat{p}_i \quad \Rightarrow \quad n(\varepsilon)=f(x),\ \ \tau(\varepsilon)=x=\sum \hat{p}_i\varepsilon_i,\quad \varepsilon_i=\begin{cases}0\\1\end{cases}$$

- $D_w^-(\alpha)$ given $\{w_i=\hat{p}_i\}$

$$\boxed{D_w^-(\alpha)\geq f(\tau)=\begin{cases}1-\tau-\sqrt{q\tau(1-\tau)}, & (1+q)\tau<1\\ \qquad 0 \qquad\qquad , & (1+q)\tau>1\end{cases}}$$



- Zone width:

$$h=\Delta\sqrt{2}=\max_\varepsilon(1-n(\varepsilon)-\tau_w(\varepsilon))\leq\begin{cases}0.5\sqrt{q}, & q\leq 1\\ q/(1+q), & q\geq 1\end{cases}$$

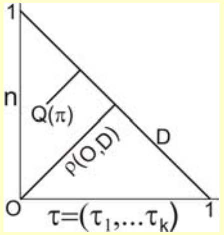
- $$\alpha\approx 97.5\%\Rightarrow h\leq 0.5\sqrt{(k+2.5\sqrt{2k})/N_-}=\begin{cases}h\leq 0.05 & k=10,\ N_-=2000\\ h\leq 0.15 & k=10,\ N_-=200\end{cases}$$

Prediction efficiency: 1-n-τ (HK score)

- HK with $\tau = \tau_w,\ w_i \neq \lambda_i / \Lambda$ say nothing about the prediction potential of π

- correct version of HK score

$$HK(\pi) = 1 - n - \sum \lambda_i \tau_i / \Lambda = 1 - n - \tau_\lambda$$
$$= \rho(Q,D) / \rho(O,D)$$



$$D = \{n + \sum \lambda_i \tau_i / \Lambda = 1\}$$

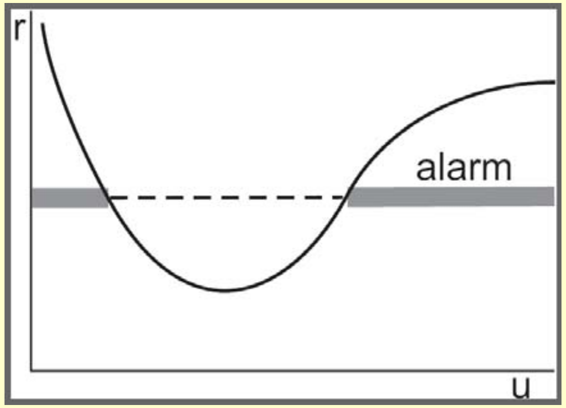
- (1-n) - τ_λ > 0, the rate of not random successes

- Optimization of HK(π)

$$\pi(t \mid G_i) = \text{alarm in } G_i \times (t, t+\Delta) \text{ as soon as } \boxed{r(t \mid G_i) > \lambda_i = \mathrm{E}r(t \mid G_i)}$$

Typical HK-values

- **Renewal** model (Molchan, 1991)



$I(t)$ is the time since the last event

$F(x) = P(t_{k+1} - t_k) < x$
interevent time distribution

$$r(t) = mF'(u)/(1 - F(u)), \quad u = I(t)$$

- If F is Weibull, Log-Normal or Gamma distribution with mean m and standard deviation σ then

σ/m	.25	.50	.75
HK score	.52-.60	.32-.38	.15-.22

↑ San Andreas

- **ETAS** model with five day lag of the data (Helmstetter & Sornette, 2003)

HK score=20%, M=6; HK score=30%, M=7

- **M8** (Kossobokov, 2005)

$$\text{HK score} = 1 - n - \tilde{\tau} = \begin{cases} 50\%, & M = 8 \\ 20\%, & M = 7.5 \end{cases}$$

$\delta n \approx \frac{1}{N} \approx 10\%$ for a single future falure-to-predict

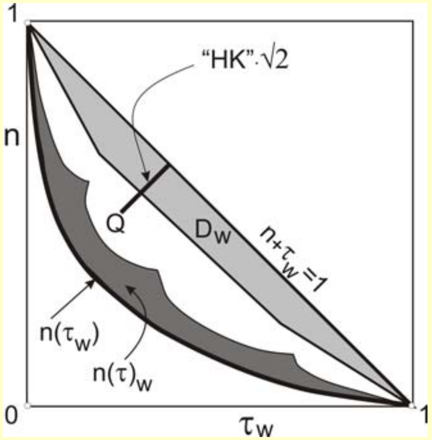
$\delta\tilde{\tau} \approx 5 - 10\%$ if M=4, 5, 6 or 7 is used to estimate the rate of M=8

Corollary:

- HK(Renewal) ≈ HK(M8) ≈ HK(ETAS)
- prediction of target events should not necessarily be based on a detailed model of seismicity

Conclusion

- be careful looking at (n,τ_w) diagram, because
  - $D_w = \{n+\tau_w = 1\}$ iff $w = (\lambda_1/\Lambda, \dots, \lambda_k/\Lambda)$
  - $\{n(\tau_w)\} = \{n(\tau)_w\}$ iff $I(t)$ is trivial
  - $dist((n,\tau_w), n+\tau_w = 1)$ say nothing about the prediction potential of π if $\{w_i\} \neq \{\lambda_i/\Lambda\}$



- $HK(\pi) = 1 - n - \sum \tau_i \lambda_i/\Lambda$, but not "$HK$"$= 1 - n - \tau_w$

  is useful as a measure of prediction efficiency at the research stage

  HK≅30%: M8 algorithm for M=8 (careful estimation)

  Renewal models for San Andreas, $\sigma/m < 0.6$ (Molchan, 1997)

- the confidence area of D_w can be useful in statistical analysis of forecasting methods

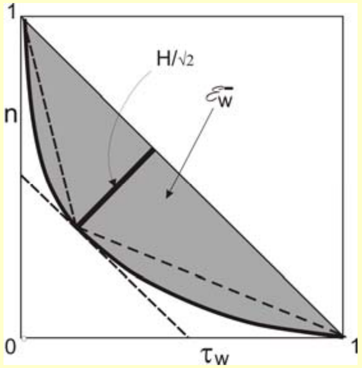
THANK YOU

- $H = \max_{\pi} HK(\pi) = \max_{\tau_\lambda}(1 - n(\tau_\lambda) - \tau_\lambda)$ **or** $A = 2 \cdot$ area of $\mathscr{E}_{\mathrm{w}}^{-}$ **?**

$|A - \hat{A}| < H(1-H)/2 < 1/8$

$\hat{A} = H(3-H)/2$

$\rightarrow$ H and the Area skill scores are equvalent

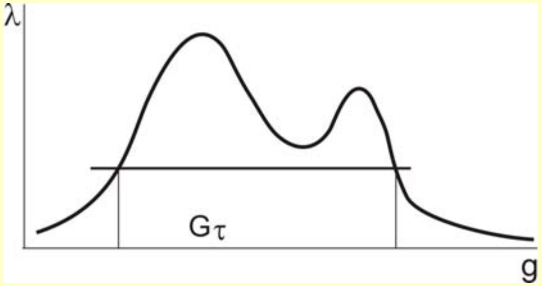


2. Distribution of $\hat{H} = max(1 - \hat{n}(\tau_\lambda) - \tau_\lambda)$: Poissonian case

$\pi(t) =$ alarm in $G_\tau \times [0,T]$

$\lambda(G_\tau)/\lambda(G) = \tau$

$N_\tau = \#\{$ success $\}$



$$\hat{H} = \underset{\tau}{max}(N_\tau / N - \tau) \overset{law}{=} D_N^+$$

$N = \#\{$ target event $\}$

D_N^+ = one sided Kolmogorov – Smirnov statistic