Regression, Mutual Information and Point Processes

David R. Brillinger
Statistics Department
University of California, Berkeley

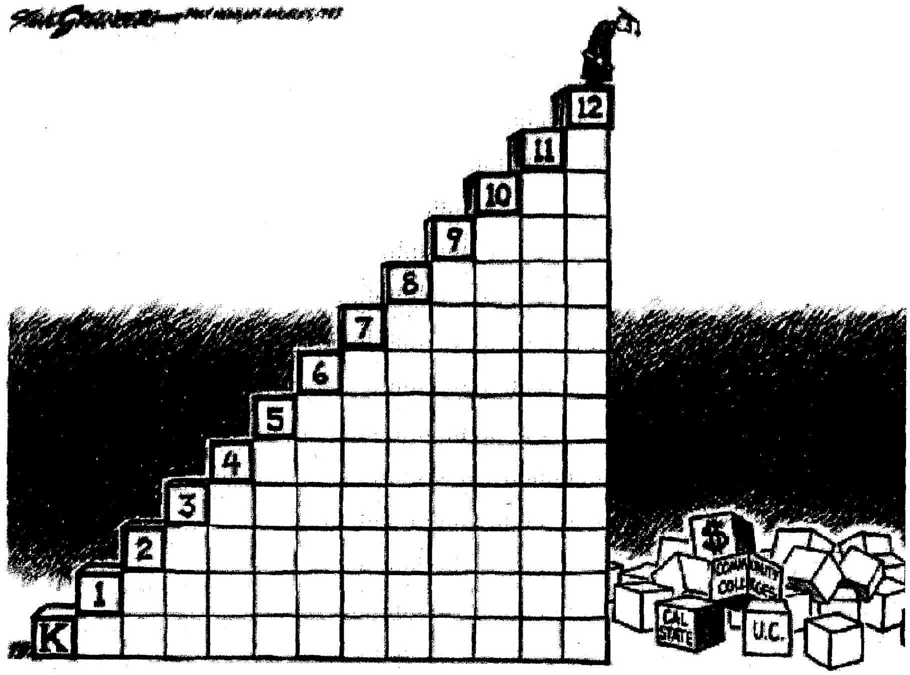
www.stat.berkeley.edu/~brill

$2\pi \neq 1$

$2\pi \neq 1$

$2\pi \neq 1$

"I hate getting all these Canadian coins, but I guess that's the price of living in Toronto."


K
1
2
3
4
5
6
7
8
9
10
11
12
$
COMMUNITY COLLEGES
CAL STATE
U.C.

Layout.

1. Introduction
2. Regression
3. Mutual information
4. Examples - neuron, soccer, river flow
5. Extensions
6. Summary
7. References
8. Some proofs

1. Introduction.

Science studies relationships

Regression analysis studies relations between variables Y and X

Question: what is the strength of a particular relationship?

One answer: the coefficient of determination

Another answer: the coefficient of mutual information

2. Classical regression.

Coefficient of determination

$$\rho_{XY}^2 = corr\{X,Y\}^2 \qquad X,\ Y \quad real-valued$$

Symmetric and invariant for:

1) independence

2) explained variation

3) **linear** dependence

4) uncertainty of estimates

One wishes for more!

3. Mutual information.

Discrete variates.

$$Prob\{X_j = j, Y_k = k\} = p_{jk}$$

$$I_{XY} = \sum_{j,k} p_{jk} \log \frac{p_{jk}}{p_j p_k}, \quad p_{jk} \neq 0$$

Continuous variates.

Given $x_j \ \varepsilon \ \delta_j$, $y_k \ \varepsilon \ \Delta_k$ and $|\delta|$, $|\Delta|$ small,

$$Prob\{X \ \varepsilon \ \delta_j, Y \ \varepsilon \ \Delta_k\} \approx p(x_j, y_k)|\delta||\Delta|$$

leading to

$$\iint p(x,y) \log \frac{p(x,y)}{p_X(x) p_Y(y)} dxdy, \qquad p \neq 0$$

Prediction - lower bound

$$E\{Y - g(X)\}^2 \geq \frac{1}{2\pi e}\exp\{2(I_{YY} - I_{XY})\}$$

Incorrect model - upper bound

$$\iint p(x,y)\log\frac{p(x,y)}{p_X(x)p_Y(y)}dxdy \geq \iint p(x,y)\log\frac{q(x,y)}{q_X(x)q_Y(y)}dxdy$$

$$\geq \underset{X,Y}{E}\left[\log\frac{q_{Y|X}(Y|X)}{q_Y(Y)}\right]$$

E.g. take q to be bivariate normal

Properties of I_{XY}.

1) Non-negative, $I_{XY} \geq 0$

2) Invariant, $I_{XY} = I_{UV}$ if 1-1 transformations

3) Measures strength of dependence

i) $I_{XY} = 0 <=> X$ indep Y

ii) $I_{XY} = \infty$ if $Y = g(X)$

iii) $I_{XZ} \leq I_{XY}$ if X indep $Z | Y$ $\quad X - Y - Z$

4) Bivariate normal, $I_{XY} = .5 * log(1 - \rho_{XY}^2)$

Disadvantage - *IM* is "just" a number

Uses.

Questions - change? trend? serial correlation?
dimension? model fit? ...

Estimation -

lag

image registration

selection of variables

model selection

association

...

Parametric estimation. Data $(x_i, y_i,\ i=1,...,n)$

Model $p(x,y\,|\,\theta)$, with X and Y independent when $\theta = 0$

$$p(x,y\,|\,0) = p_X(x)p_Y(y)$$

Test of independence of X and Y

$$\hat{I}_{XY} = -\log(\textit{likelihood ratio})/n$$

(n sample size)

Approximate null distribution, when (X_i, Y_i) independent

$$\chi^2_\nu \,/\, 2n, \quad \nu = dim(\theta)$$

$$\iint p(x,y) \log p(x,y)dxdy - \iint p(x)p(y) \log p(x)p(y)\ dxdy$$

Non-parametric estimation.

$\hat{p}(x,y)$ estimate of $p(x,y)$, e.g. histogram or kernel-based

$$I_{XY} \hat{=} \sum_{j,k} \hat{p}(u_j,v_k) \log \frac{\hat{p}(u_j,v_k)}{\hat{p}_X(u_j)\hat{p}_Y(v_k)}$$

Approx null distribution

$$\chi_\nu^2 \, / \, 2n, \quad \nu = (J-1)*(K-1)$$

Non-null

$$\hat{I}_{XY} \rightarrow I_{XY} \ in\ prob,\ etc.$$

(Entropy - Bilmes, Fernandes, Hall & Morton, Joe, Kozachenko & Leovenko, Parzen, Robinson)

The point process case. Isolated points

Univariate. points $\{\tau_k\}$, counts $N(t) = \#\{\tau_k \le t\}$,

intervals $\{Y_k = \tau_{k+1} - \tau_k\}$, history $H_N^t = B(\tau_k \le t)$

conditional intensity

$$Prob\{dN(t) = 1 \mid H_N^t\} \approx \mu_N(t)dt$$

likelihood

$$\prod_k \mu_N(\tau_k) \exp\{-\int_0^T \mu_N(t)dt\}$$

entropy

$$E_N \{\int_0^T \log \mu_N(t)\, dN(t) - \int_0^T \mu_N(t)dt\}$$

Bivariate. $M(t) = \#\{\sigma_j \le t\}$, $N(t) = \#\{\tau_k \le t\}$

history $H_{MN}^t = B(\sigma_j,\ \tau_k \le t)$

$$Prob\{dM(t) = 1 \mid H_{MN}^t\} \approx \gamma_M(t)dt, \quad Prob\{dN(t) = 1 \mid H_{MN}^t\} \approx \gamma_N(t)dt$$

entropy

$$\underset{M,N}{E}\ \{\int_0^T \log\,\gamma_M(t)dM(t) + \int_0^T \log\,\gamma_N(t)dN(t) - \int_0^T \gamma_M(t)dt - \int_0^T \gamma_N(t)dt\}$$

mutual information

$$\underset{M,N}{E}\ \{\int_0^T \log\frac{\gamma_M(t)\gamma_N(t)}{\mu_M(t)\mu_N(t)}dt\}$$

Aplysia californica

Examples.

a) Interval analysis - Aplysia californica

Neuron, $L10$, discharging spontaneously

Spike train $\{\tau_k\}$

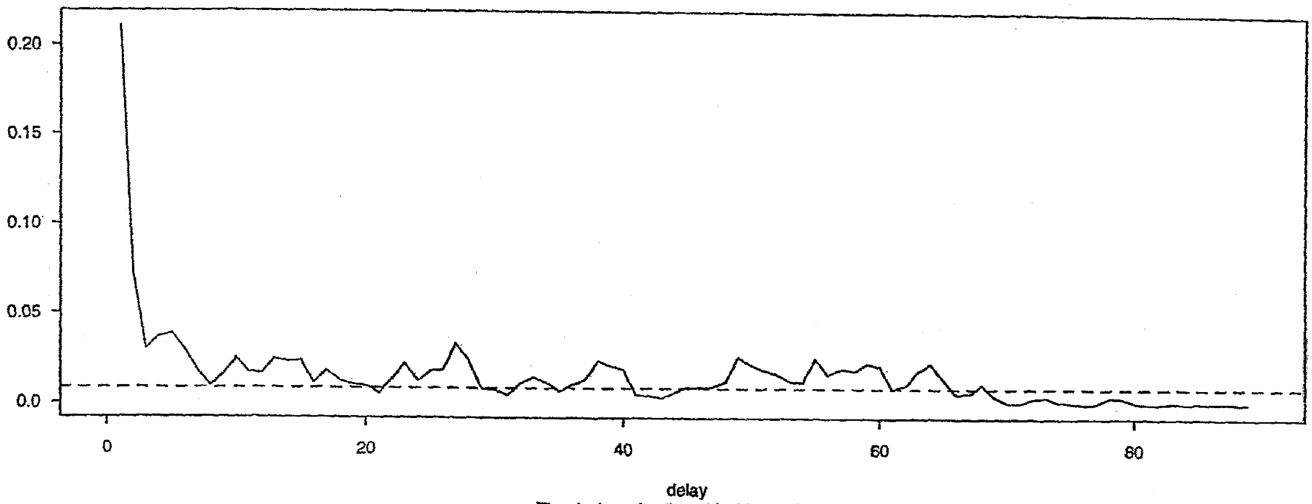
$\{Y_k = \tau_{k+1} - \tau_k\}$ intervals between firings

Many models imply that intervals i.i.d. (renewal)

Estimate coefficient of determination and MI, as functions of lag, k

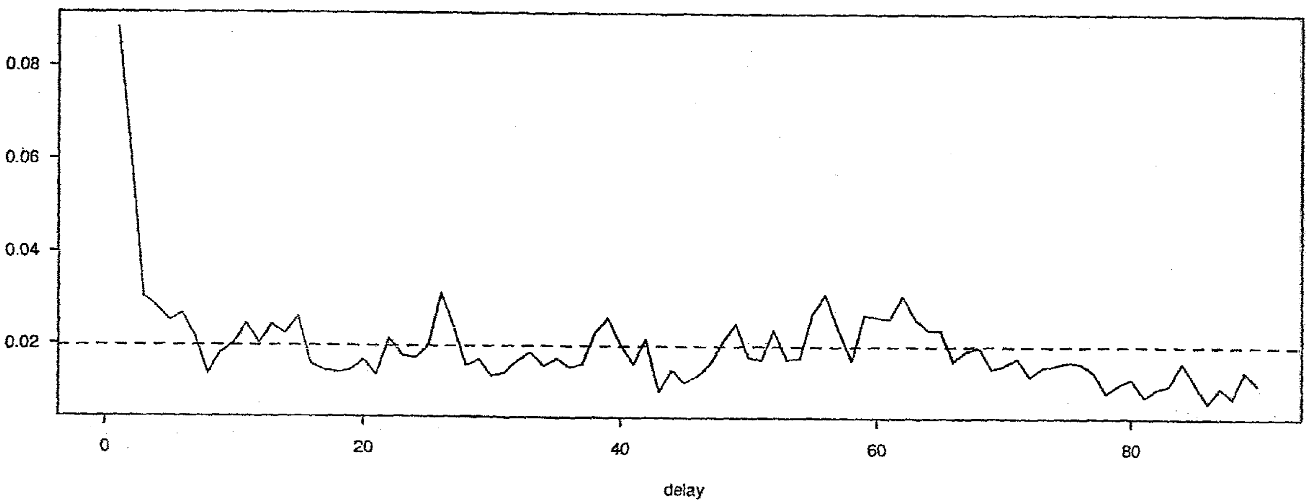
Intervals between firings of an Aplysia neuron

Coefficient of determination

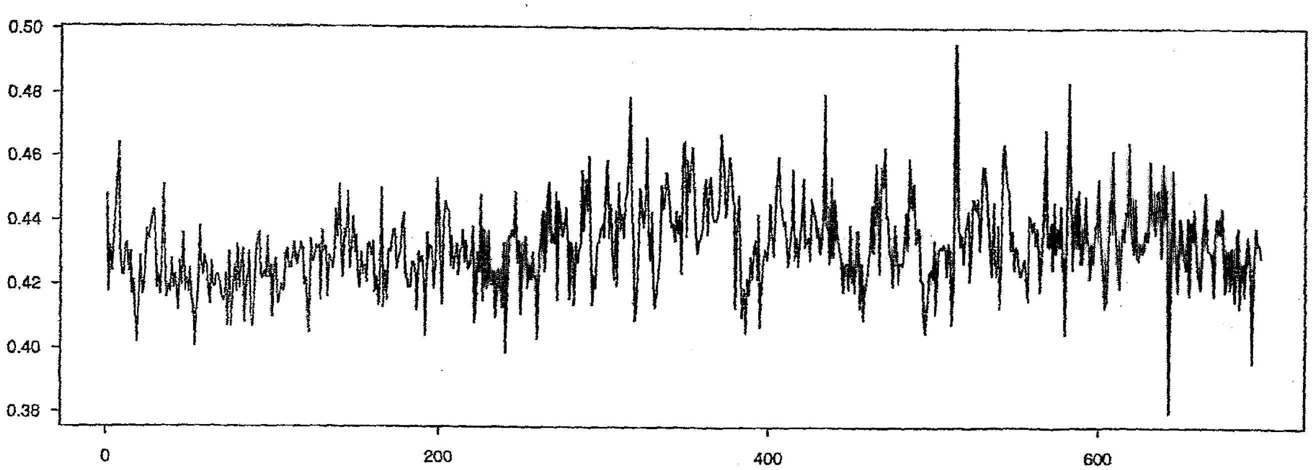


The dashes give the critical level of 99%

Mutual information



The data



We estimate the 99% critical level via random permutations of the intervals.

There is evidence against the assumption of a renewal process

Unusual test of normality

The world's first game of soccer...
AHAJOKES.COM

b. Two discrete variables - soccer

Question - In which country is the relationship strongest between the number of goals teams score and and their playing at home?

Data for the Premier Leagues, 2001-2002

http://sunsite.tut.fi/rec/riku/soccer2.html

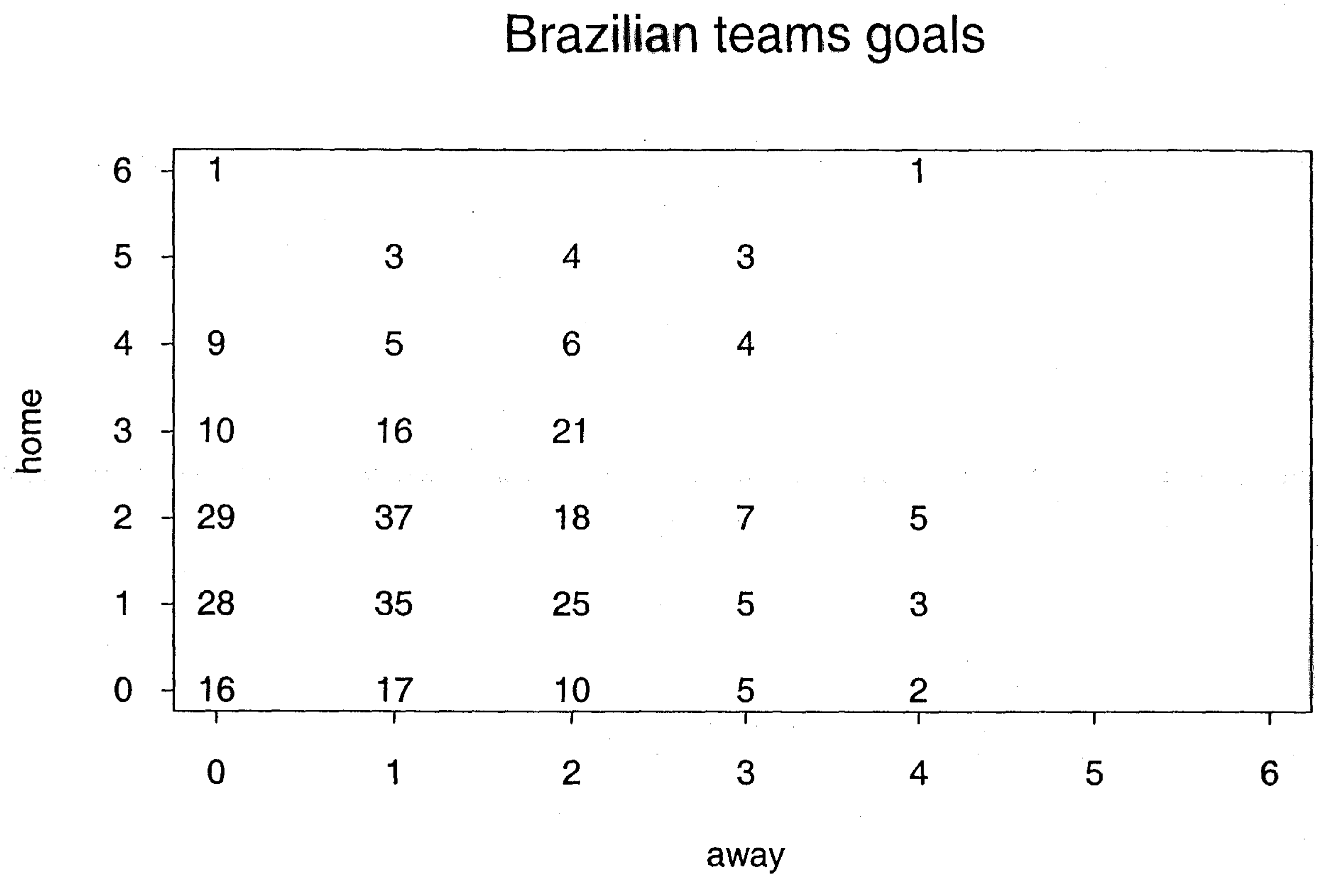
Y = 0,1,2,3,4+ goals

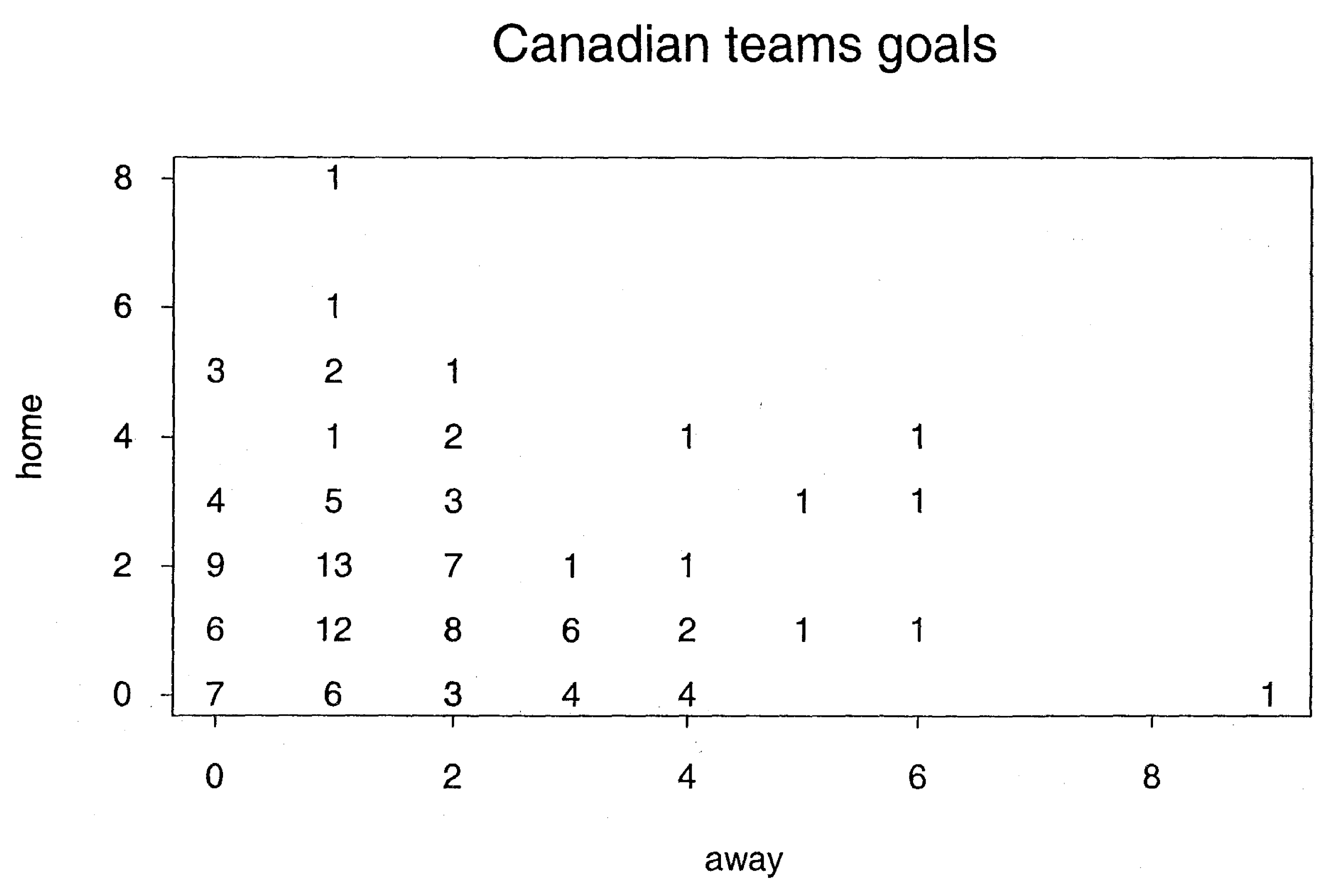
X=1,0 - playing at home or away

Brazilian Serie A - Games played so far

Aug 10, 2002	Vasco	- Figueirense	2 - 0
	Sao_Paulo	- Paysandu	4 - 2
	Parana	- Sao_Caetano	2 - 1
	Santos	- Botafogo-RJ	2 - 1
	Goias	- Portuguesa	3 - 1
Aug 11, 2002	Fluminense	- Cruzeiro	5 - 1
	Palmeiras	- Gremio	1 - 1
	At.Mineiro	- Corinthians	1 - 2
	Internacional	- Flamengo	1 - 3
	Guarani	- Atletico-PR	2 - 1
	Coritiba	- Vitoria	1 - 0
	Bahia	- Gama	1 - 0
	Juventude	- Ponte_Preta	1 - 0

Aug 14, 2002	Sao_Caetano	- Fluminense	2 - 0
	Botafogo-RJ	- At.Mineiro	1 - 1
	Cruzeiro	- Palmeiras	1 - 1
	Gremio	- Vasco	[illegible]

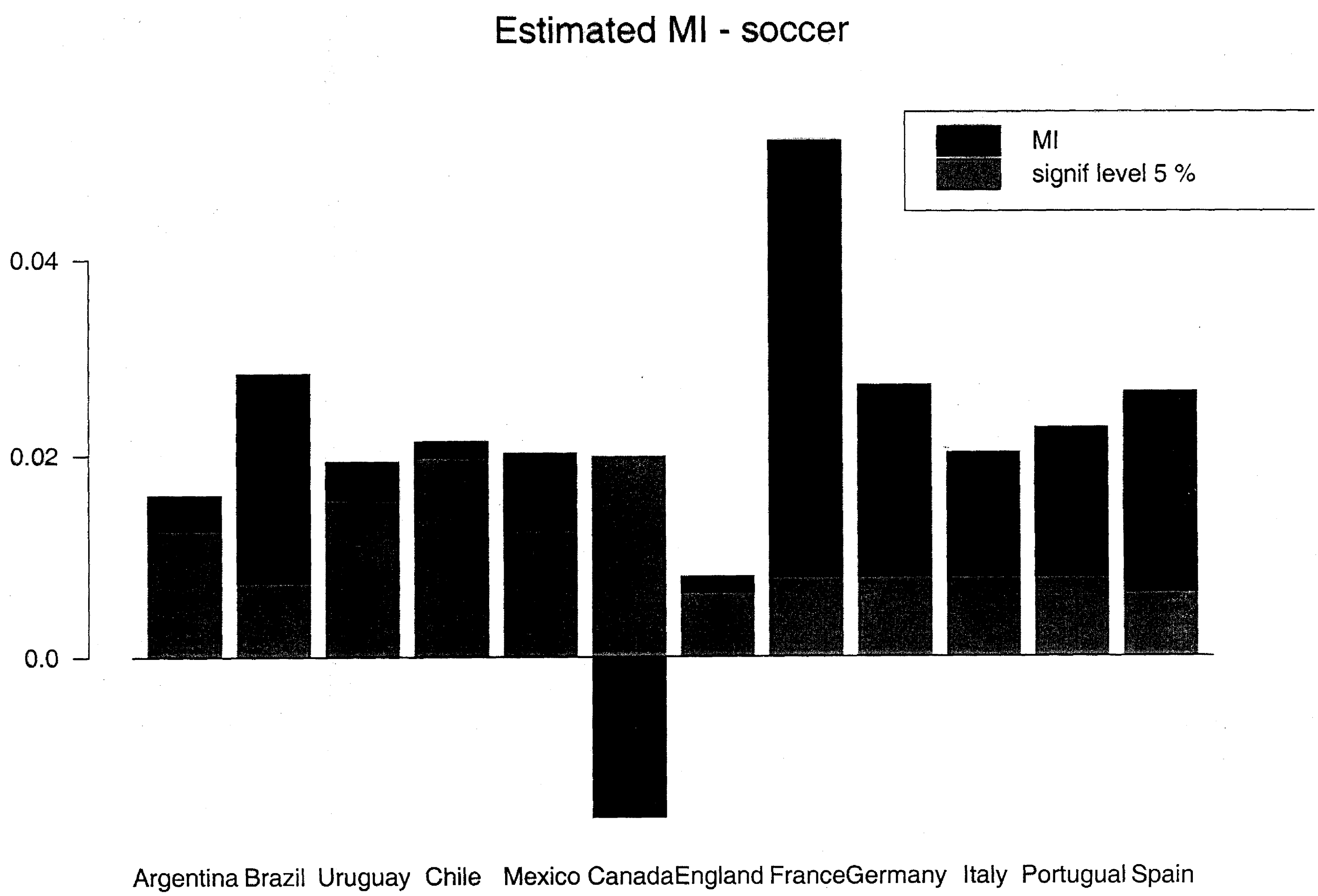

Brazilian teams goals
home
away
0
1
2
3
4
5
6


Canadian teams goals
home
away
0
2
4
6
8

Discussion.

Marginal analyses

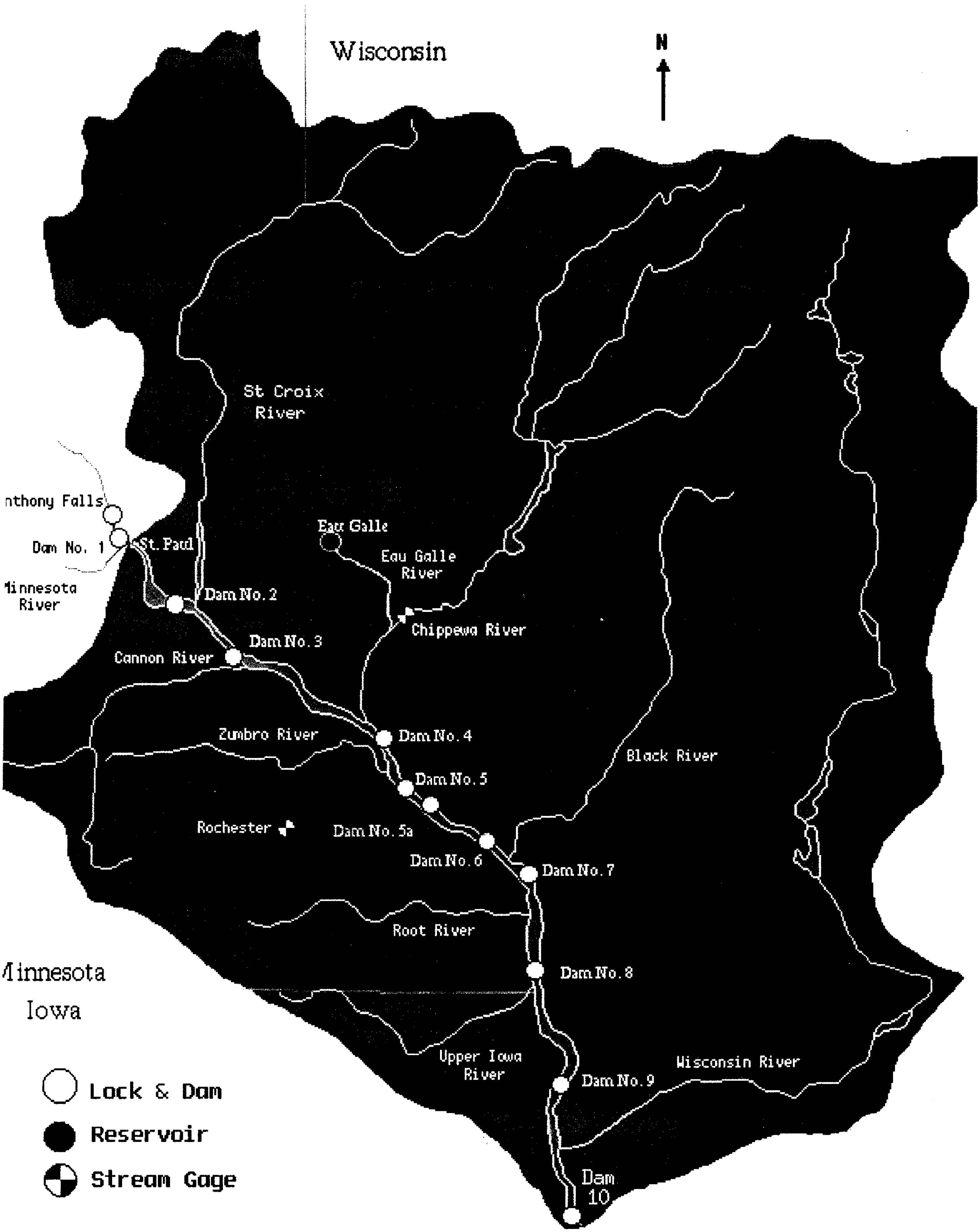
Games i.i.d. ?



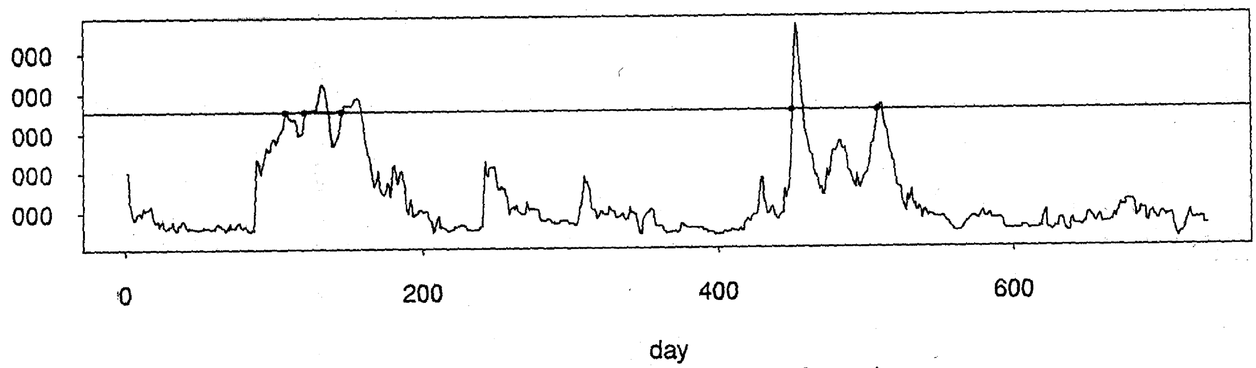
Estimated MI - soccer
MI
signif level 5 %
0.04
0.02
0.0
Argentina
Brazil
Uruguay
Chile
Mexico
Canada
England
France
Germany
Italy
Portugual
Spain

The Great Mississippi River Flood of 1927, photographed in Illinois on March 25.

"The Floods of 1927 in the Mississippi Basin," Frankenfeld, H.C., 1927 Monthly Weather Review Supplement No. 29

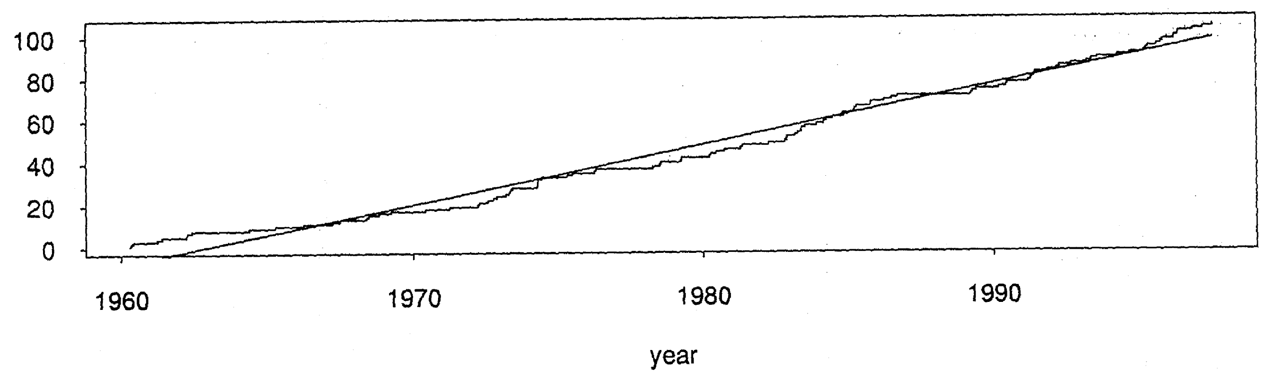

Wisconsin
N
St Croix
River
nthony Falls
Dam No. 1
St. Paul
Minnesota
River
Eau Galle
Eau Galle
River
Dam No. 2
Chippewa River
Dam No. 3
Cannon River
Zumbro River
Dam No. 4
Black River
Dam No. 5
Rochester
Dam No. 5a
Dam No. 6
Dam No. 7
Root River
Dam No. 8
Minnesota
Iowa
Upper Iowa
River
Wisconsin River
Dam No. 9
Dam
10
Lock & Dam
Reservoir
Stream Gage

Dam 8 flow in 1960 & 1961



threshold at 10th percentile of yearly maxima

Cumulative count of upcrossings



c. Bivariate point process case. Mississippi River flow

Question - Strength of dependence between flows as function of distance between dams?

10 dams from St. Paul, Minnesota to Iowa

Daily flow from 1/1/1960 to 31/12/1997

$\{Y_i(t), t=0,\pm1,\pm2,\ldots, i=1,\ldots,10\}$

Have distances between dams

Upcrossing times of 10th percentile of series yearly maxima

Model.

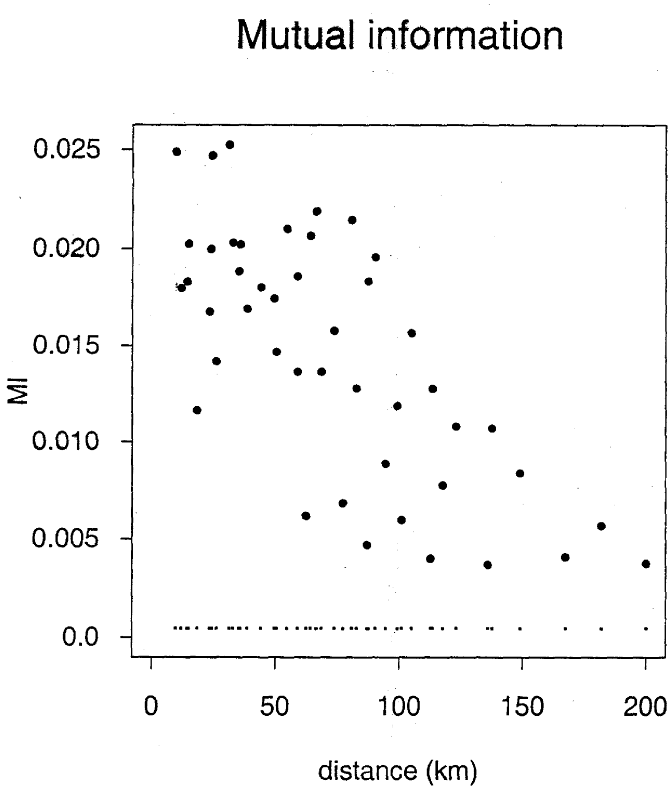
Two 0-1 series X_t and Y_t.

$$logit\ Prob\{X_t=1\,|\,H_t\} = \alpha_1 X_{t-1}+\ldots+\alpha_p X_{t-p} + \beta_1 Y_{t-1}+\ldots+\beta_q Y_{t-q}$$

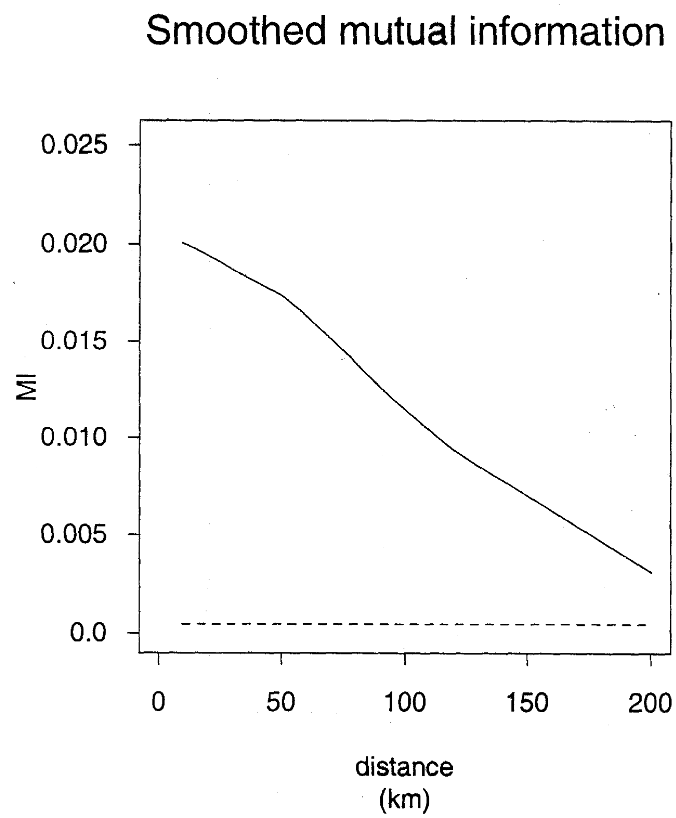
$$logit\ Prob\{Y_t=1\,|\,H_t\} = \gamma_1 Y_{t-1}+\ldots+\gamma_p Y_{t-p} + \delta_1 X_{t-1}+\ldots+\delta_q X_{t-q}$$

Include recent times and some from a year earlier

Proceed via likelihood ratio test of no connection.



Mutual information
0.025
0.020
0.015
0.010
0.005
0.0
MI
0
50
100
150
200
distance (km)



Smoothed mutual information
0.025
0.020
0.015
0.010
0.005
0.0
MI
0
50
100
150
200
distance
(km)

Discussion.

For asymptotics need full model to "fit".

Estimate spectral density matrix of residuals.

5. Extensions.

MI decomposes , $I_{YX} = I_{YX_1} + I_{YX_2 | X_1}$

Uncertainty (permutations, jackknife, bootstrap,...)

Local estimates

Multivariate cases

Other estimates of entropy.

Discrete alphabet

R_n: time until the first n-string repeats

$$(\log R_n)/n \to \textit{entropy} \quad a.s.$$

6. Summary.

MI is a concept extending correlation, substitutes for r^2 and R^2

"The hypothesis of independence is rejected." becomes

"The estimated strength of dependence is $\hat{MI}$*."*

Functional forms useful, new parameters suggested

Google: *corr* 3,340,000 hits, *MI* 40,400 (5/31/2003)

7. References

Brillinger, D. R. (2002). Second-order moments and mutual information in the analysis of time series. Pp. 64-76 in Recent Advances in Statistical Methods (Ed. Y. P. Chaubey). Imperial College Press, London.

Cover, T. and Thomas, J. (1991). Elements of Information Theory. J. Wiley, New York.

Granger, C. W. J. and Lin, J-L. (1994). Using the mutual information coefficient to identify lags in nonlinear models. J. Time Series Anal. 15, 371-384.

Hall, P. and Morton, S. C. (1993). On the estimation of entropy. Ann. Inst. Statist. Math. 45, 69-88.

Joe, H. (1989). Relative entropy measures of multivariate dependence. J. American Statistical Association 84, 157-164.

Li, W. (1990). Mutual information functions versus correlation functions. J. Statistical Physics 60, 823-837.

8. Some proofs.

Jensen's Inequality.

For g convex

$$E\{g(Y)\} \geq g(E\{Y\})$$

with equality iff Y is degenerate