15 Sept 2002

Inferential and ~~Dese~~ Descriptive Statistical Analysis of Data from Microarray Experiments.

A: Which genes ~~should~~ show changed level of expression
SAM; Baysian Analysis.

B: How to represent results of time-course experiment
clusty - Dendrogram.

(1) Hypothesis testing: Control vs Treatment style hypothesis
Null hypothesis, type 1 and type 2 errors.
False positives, false negatives.
p values: Reduces false positives, increases false negatives.

(2) Testing multiple hypothesis — Higher # false positives.

Alternative methods: SAM
. BAYSIAN ANALYSIS.

~~**~~ ~~city~~

Data: U1A U1B I1A I1B U2A U2B I2A I2B.
Effects of Radiation: Lymphoblastoid cells:
oligonucleotide arrays: Diff between pm & mm
Spot - log ratio of Red/Green.

Notation: M_{ij} why not fold increase? ——— may false positives 5/1, 50/10, 500/100

$$Z_i = \frac{\bar{M}_{iI} - \bar{M}_{iU}}{a + S_i}$$

——→ More variation at low value, signal/noise ratio is low

a = lowest ~~variance~~ cell. var.
= 90th percentile.

$Z_i = \frac{m}{\quad}$

P.T.O

②

The heart of the methods is distribution of 'null' scores when there are no differences.

How to compute:

U1A − U1B, I1A − I1B, U2A − U2B, I2A − I2B

Alternatively — Make two groups with 2 U and 2 I & take middle of means ⟶ 36 values.

Random permutation of sym...

Obtain d_{O_i} d_{E_i} null

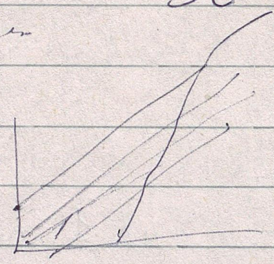


Fig 2 of Goss

False positives: = Average no's over 36 perm. of the d^S which fall in this region.

Table 1 of Goss.

Comparison of fold method shows that SAM is better

Fig 4 of Goss.

Bayesian Approach

$$f(z) = p_0 f_0(z) + (1-p_0) f_1(z)$$

Estimate by logistic regression

$$p_1 > 0 \Rightarrow 1 - p_0 \frac{f_0(z)}{f(z)} > 0$$

$$p_1(z) = 1 - \frac{p_0 f_0(z)}{f(z)}$$

$$1 > \frac{p_0 f_0}{f(z)}$$

≈20 replicated z, scores, z_0, $\pi(z) = \frac{f(z)}{f(z) + B f_0(z)}$

$$1 > (1-p_1)\frac{f_0}{f}$$

$$\frac{f_z}{f_0} > 1 - p_1$$

$$1 - \pi = \frac{B f_0(z)}{f(z) + B f_0(z)}, \quad \frac{1-\pi}{\pi} = \frac{B f_0(z)}{f(z)}$$

$$p_1 > 1 - \frac{f_z}{f_0}$$

25 sept 2002

$P_1 + P_0 = 1$

$$P_1(Z) = 1 - \frac{P_0 f_0(Z)}{f(Z)} \qquad f(Z) = P_1 f_1 + P_0 f_0$$

$$1 - \frac{P_0 f_0}{f} > 0 \implies 1 > P_0 \frac{f_0}{f}$$

$$> (1-P_1) f_0/f$$

$$\implies f > f_0 - P_1 f_0 \quad f/f_0 > 1 - P_1 \quad P_1 > 1 - \frac{f}{f_0}$$

Smallest $\frac{f(Z)}{f(0)}$ Compute $\frac{f}{f_0}$ by logistic regression

Randomized f(Z)
f(Z) is normal, so f_0 compute, so f_1
Compute P_0
Compute $P(Z)$. — Threshold.
Simulate Enhancers with random d^s — false

Best Threshold of P & obtain spec genes
Validate by null set

1. Eisen PNAS Dec. 1998 15 Sept 2002

Cluster ~~analysis~~ analysis . dendrograms

(i) Visualization of time course (No inference).

(ii) Dendrograms drags, weighs.

(iii) Tests of dendrogram ——— randomization

Randomization / simulation as a powerful ~~tool~~ method to explore the robustness / reliability of the conclusions.

Fibroblasts stimulated with serum.

my log ——— 12 time pts. 0, 15', 30', 1, 2, 3, 4, 8, 12, 16, 20, 24

Red / green : −3 green +3 red. , 8000-9000 genes.

(iv) Many conditions ——— 'same' genes are clustered together.

~~F~~ Genes selected if >3 fold, in at least 2 conditions.

16 Sept. 2002

Intro of. → T1

Outline. → T2, Refs 3

1 Hyp Tests → T4

2 Many hyp. → T5

(1) SAM → Z, Z

SAM plus
False positive control, Rel to Table

Contributors.

(2) Bayes approach:

$f(z) = p_0 f_0(z) + p_1 f_1(z)$

$p_1(z) = 1 - p_0 \frac{f_0(z)}{f(z)}$.

Comments → Bayes regimes

(3) Maximum Likelihood Lee et al.

2 april 2003

suse 6.3 linux

nvjah → 725476 K.

---

du for d disc → 786896 ☆

CAMBRINGE → 686336 → D 686340

---

disk d hdb1/ - 539980 ☆ ddwll

disc f hdb5/ ~~798582~~ bytes.
10~~879~~ files

10879 files 516 594 225 bytes
1667 directories.
D disk

F
disc f, /hdb5/ 715680 → 585680 on des server

Computations & [illegible] 1 April 2003

[illegible] [illegible]

[illegible]

what is need, just 1 number?

Dill. val d x min ?

How many [illegible].

[illegible], eg [illegible]

Do you already have agree [illegible] 1800?

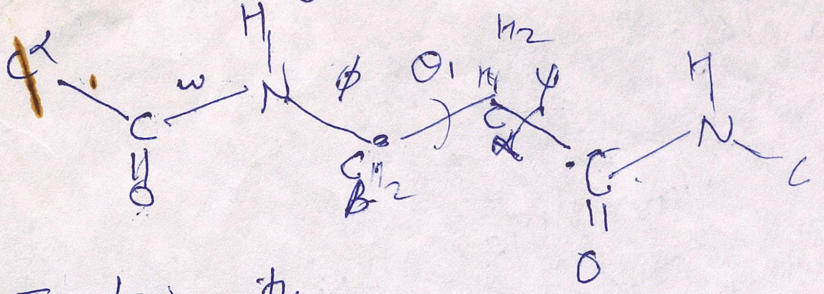
is it 3 [illegible] [illegible] 250 & 3 [illegible] [illegible] 500

what [illegible]

Smin = Smin / 5

and smin = 25

Hydrogen bonding schemes for β glycine ele. 29 June 2009



To begin with, 6 atoms per unit

CA, C, O, N, H, CB.,

Fix ~~CA~~ C at origin, O along -ve y, CA & CN in bond angl

Cβ using ω, H using ω.

For next unit, C^{α}_{2} using C^{β}, N H, ϕ_1

C_2 using C^{α}_{2}, $C\beta_1$, N_1, θ_1 = 180

N_2 using C_2, $C\alpha_2$, $C\beta_1$, ψ.

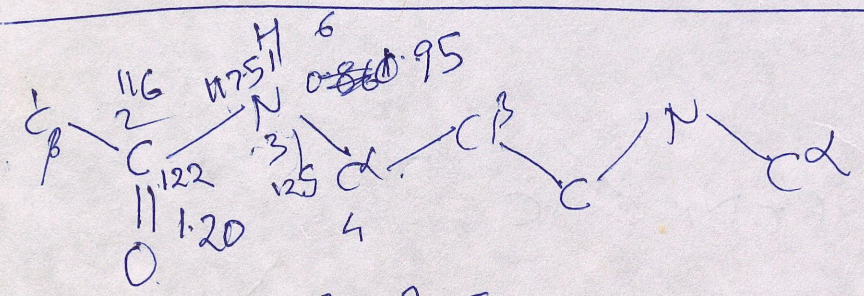
O_2 (same as above), ψ + 180

H_2 using N_2, C_2, ... ω

$C\beta_2$ using N_2 C_2 ... ω.

---

6 JULY



5   4-3-2, 5

~~4-3-2~~, 1

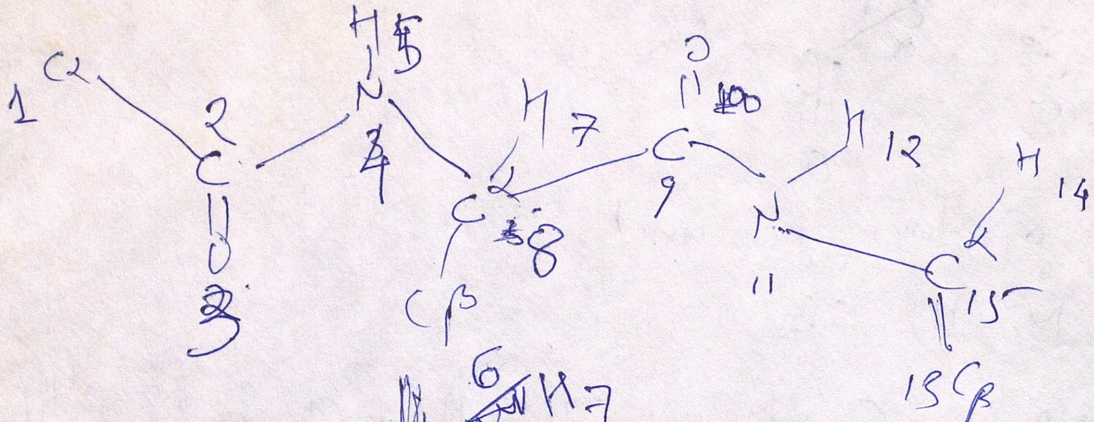
~~4-3~~

1-2-3-6

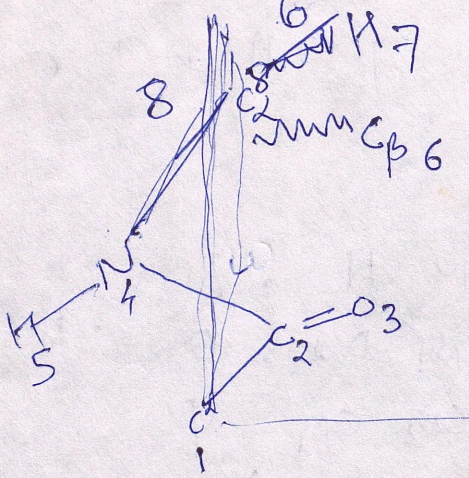
HB, ON dist 3.7-2.6

ces machine   pbjan2k4   5 July

ramplot2.f , check numbers



coord



mbl 4,11,2,1,-2,-1

mbq 6,5,-1,3,3,2,3

mta 7,7,-5,7,2,4,5

Fix =O with C_{α}, N, C, & 0° cis

H on N : with ..., C, N & 180 (trans)

φth atom = C, ,

PHI = C–N,–α–C

C^{β} = 6 = connected to 8 , N, , φ+240

1, N; , C^{α},C,N ; ψ = N,Cα,C,N

Defn of unit

C_{β} – C(=O) – N – C^{α} – C_{β}