Rényi continuous entropy of DNA sequences

Susana Vinga, Jonas S Almeida

J. Theor. Biol. **2004** vol. 231(3) pp. 377-388

Current version: 2004-09-20 (v1.1) from file renyi1v1.zip

1. Introduction

This document briefly exemplifies the use of some MATLAB functions for the calculations of Rényi quadratic entropy of DNA sequences.
For further information and updates visit http://bioinformatics.musc.edu/renyi
Contact: svinga@itqb.unl.pt

2. Example

Tested on MATLAB© version 6 – release 13.
The following output can be obtained by running script `example.m`.

COMMANDS	COMMENTS
`>>seqfile='MC0.seq';` `>>nrep=50;`	file name number of replicas

`>>s=readfasta(seqfile)`
 	
`s =`
 	
`title: 'source file: MC0.seq'`
`legend: {'MC0'}`
`sequence: {[1x2000 char]}`
	create struct variable `s` with information about sequence; read from file '`MC0.seq`' (in FASTA format)
	
	alternatively, do:
	`>>load MC0`

`>>c=usm_make(s.sequence{1});`	extract CGR/USM coordinates `c`
`>>load sig2;`	load variable with vector `sig2`; corresponds to kernel variances `sig2=10.^[-10:0.25:2]`

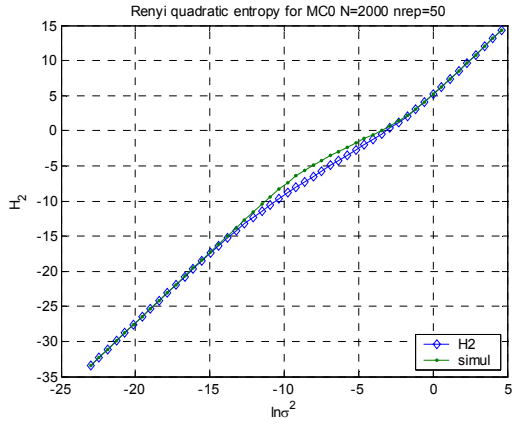
`>>H2=renyi2usm_fast(c,sig2,s.legend{1});`
 	
`sig2 = 1e-010`
`sig2 = 1.7783e-010`
`sig2 = 3.1623e-010`
`sig2 = 5.6234e-010`
`sig2 = 1e-009`
`sig2 = 1.7783e-009`
`sig2 = 3.1623e-009`
`sig2 = 5.6234e-009`
`sig2 = 1e-008`
`sig2 = 1.7783e-008`
`sig2 = 3.1623e-008`
`sig2 = 5.6234e-008`
`sig2 = 1e-007`
`sig2 = 1.7783e-007`
`sig2 = 3.1623e-007`
`sig2 = 5.6234e-007`
`sig2 = 1e-006`
`sig2 = 1.7783e-006`
	calculate Rényi continuous quadratic entropy based on CGR/USM coordinates `c` and gaussian kernel variances `sig2`; save results to filename *`MC0`*
	
	…run through all `sig2` values…

sig2 = 3.1623e-006 sig2 = 5.6234e-006 sig2 = 1e-005 sig2 = 1.7783e-005 sig2 = 3.1623e-005 sig2 = 5.6234e-005 sig2 = 0.0001 sig2 = 0.00017783 sig2 = 0.00031623 sig2 = 0.00056234 sig2 = 0.001 sig2 = 0.0017783 sig2 = 0.0031623 sig2 = 0.0056234 sig2 = 0.01 sig2 = 0.017783 sig2 = 0.031623 sig2 = 0.056234 sig2 = 0.1 sig2 = 0.17783 sig2 = 0.31623 sig2 = 0.56234 sig2 = 1 sig2 = 1.7783 sig2 = 3.1623 sig2 = 5.6234 sig2 = 10 sig2 = 17.7828 sig2 = 31.6228 sig2 = 56.2341 sig2 = 100 ##Renyi2 file: renyi2usm_MC0.rn2	…and save file with results ‘renyi2usm_MC0.rn2’; extension *.rn2
>>N=length(s.sequence{1});	
>>disp('Begin simulations...'); >>SIM=simul_renyi_usm2(N,sig2,nrep); >>disp('...simulations ended'); >>med=median(SIM,1);	Begin simulations of random sequences (with same length N) ->number of replicas= nrep (verification: if already saved simply read from *.sim file)
Begin simulations... ***Simulation is not saved. Running 50 replicas for length N=2000 Replica nb... ...1 sig2 = 1e-010 […] sig2 = 100 ##Renyi2 file: renyi2usm_AUX.rn2 ...2 […] ...50 sig2 = 1e-010 […] sig2 = 100 ##Renyi2 file: renyi2usm_AUX.rn2 ### FILE created/read: N2000R50.sim ...simulations ended	repeat calculations…
>>figure; >>plot(log(sig2),H2,'d-',log(sig2),med','.-'); >>title(['Renyi quadratic entropy for ' s.legend{1} ' N=' num2str(N) ' nrep=' num2str(nrep')]); >>xlabel('ln\sigma^2');ylabel('H_2');gri	plots results of Rényi entropy H2 as a function of ln(sig2)

d; >>legend({'H2' 'simul'},4); >>saveas(gcf,['H2' s.legend{1} '.fig'],'fig'); ### FILE created: H2MC0.fig	saves figure
>>step=log(10)*0.25 step = 0.5756	corresponds to elements of >>step=diff(log(sig2))
>>figure >>plot(log(sig2(1:end- 1)),diff(H2)/step,'d-',log(sig2(1:end- 1)),diff(med)'/step,'.-'); >>title(['Derivative of Renyi quadratic entropy for ' s.legend{1} ' N=' num2str(N) ' nrep=' num2str(nrep')]); >>xlabel('ln\sigma^2');ylabel('\DeltaH_2 ');grid;	plots results of Rényi derivative ΔH2 as a function of ln(sig2)
>>legend({'H2' 'simul'},4); >>saveas(gcf,['dH2' s.legend{1} '.fig'],'fig'); ### FILE created: dH2MC0.fig	saves figure

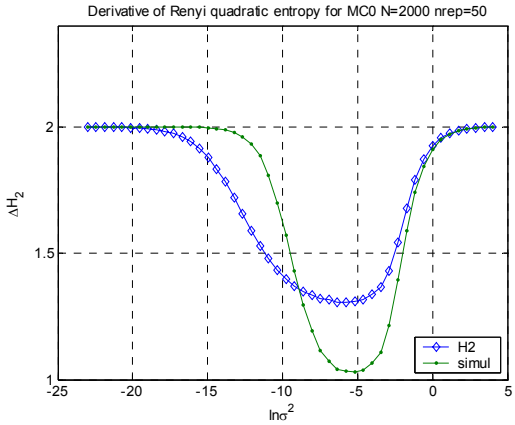
Graphs obtained:

H2MC0.fig



Renyi continuous quadratic entropy of sequence **MC0** (length N=2000) as a function of the Gaussian kernel s used in Parzen's Method (H2). Comparison with nrep=50 random sequences of the same length, obtained by Montecarlo simulation (simul) – graph represents the median values obtained.

dH2MC0.fig



Discrete derivative of H2 for sequence **MC0** – corresponds to slope of previous graph.