Frequently Asked Questions (FAQ) for bionet.info-theory Biological Information Theory and Chowder Society

version = 2.29 of bionet.info-theory.faq.html 2001 August 28

http://www.ccrnp.ncifcrf.gov/~toms/bionet.info-theory.faq.html

---

Summary:

This is the Frequently Asked Questions monthly posting for BITCS. The news group bionet.info-theory is a forum for discussing information theory in biology and for tossing food for thought around. Other interesting mathematical problems in biology are also welcome, as we will try our best to take the log of them, so as to convert them into information theory problems.

*** NEWCOMERS PLEASE NOTE:

Although the name of this group, bionet.info-theory has the word "info" in it, this newsgroup is NOT an appropriate forum for persons seeking information about general questions related to biology or medicine! This newsgroup is devoted to DISCUSSIONS ABOUT BIOLOGICAL APPLICATIONS OF INFORMATION THEORY, principally referring to Shannon's theory of information, although we also discuss the mathematical and physical meaning of entropy, alternative definitions of information, and related fundamental issues in information theory and biology.

- **Questions about The BITCS, the newsgroup, and this FAQ**
  - What is The Biological Information Theory and Chowder Society?
  - How Do I obtain bionet.info-theory BY EMAIL?
  - Where Did I Get This FAQ File Originally?
  - What is the IP number of the FAQ archive?
  - Where Are the Bionet Archives?
  - Are There Other Archives?
  - I Posted But Nothing Happened?!?
  - What is an Appropriate Posting?
  - What Can I Do About Inappropriate Postings?
  - Should I send private email to someone to respond to a posting or to ask a question?
  - What is the official word on copyright of this FAQ?

- Who Takes Care of This Group?
- What Kind of Questions Are Appropriate For Discussion?
- When and Where are Meetings?
- Acknowledgments

- **Questions about Information Theory**

  - What is Information Theory?
  - Is There a Quick Introduction to Information Theory Somewhere?
  - I'm Confused: How Could Information Equal Entropy?
  - How Can I Learn More About Information Theory and Biology? References
    - REFERENCES - General
    - REFERENCES - Information Theory
    - REFERENCES - Jaynes
    - REFERENCES - Schneider
    - REFERENCES - Yockey
    - REFERENCES - Adleman and papers related to molecular computation
    - REFERENCES - Gad Yagil and papers related to Algorithmic Information Theory (AIT) or Algorithmic Complexity
    - REFERENCES - Entropy on the World Wide Web
  - Will Authors Send Me Papers?
  - Where Can I Get BIG Coins?
  - Are there other organizations for information theory?
  - What are Sequence Logos?
    - How Do I find Sequence Logos on the Web?
    - Is There a Shell Script for Making Sequence Logos?
    - Is There a World Wide Web Page for Making Sequence Logos?

What is The Biological Information Theory and Chowder Society?

The Biological Information Theory and Chowder Society (BITCS) is a group of scientists interested in the biological applications of information theory (thus the "BIT") who meet informally for dinner (thus the "CS") from time to time in the Washington, DC, area. At our dinners we have only one rule --- food fights are discouraged.

The guys who started this thing did it because we weren't certain we understood the biological implications of information theory. Some of us are more comfortable with the mathematical machinery and assemble biological systems into grand canonical ensembles whether they want to be there or not; and some of us think they understand what the biological systems are doing but can't take a log to base 2. What we try to do is pry from one another the bits of knowledge that will help us understand what's going on.

Some of the topics up for discussion in our group are:

- biological applications of information theory
- biochemical molecular machines and computers
- computer methods for recognition of molecular structure and function

- database organization for biomolecular information
- nanotechnology
- the limits of computation
- "dissipation-less"(?) computation
- Maxwell's demon
- anecdotes and humor about all these topics
- methods and theories of molecular computation
- macroscopic versus microscopic thermodynamics

A few relevant papers are given in the references.

The group started when Tom Schneider was introduced to John Spouge in 1988. Tom bounced his ideas about molecular machines off John, and John kept finding flaws. Tom would go away rather unhappily for a month and then find a solution. But John was always one step ahead... (and still is, on last account.) Tom gave a talk about molecular machines at the Lambda Lunch meeting on the Bethesda NIH campus, and John introduced John (Steve) Garavelli. We all got together with Peter Basser for dinner once in a while to talk about information theory. Steve brought in one of the first people to apply information theory to biology, Hubert Yockey. Steve Garavelli dubbed the group the "Biological Information Theory and Chowder Society", which it is still called. We are known sometimes as 'chowderheads', and talk about food fights, but so far have only had electronic food fights! We hold dinners in Bethesda, Maryland on random occasions.

When our informal mailing list became difficult to handle, we petitioned to start a bionet news group. We have held roaring discussions and look forward to more, and everyone is welcome to join. You can look at some of the ancient discussions in the bionet archives. If you are uncertain about something, quit lurking and ask on the net. It may well be that what bothered you is the key to a new piece of information theory in biology. (The major advances so far have been by things that REALLY bugged people.)

We will also announce when and where our (irregular) eatings are and you are welcome to join if the travel is not too far. John Spouge usually makes the arrangements. If you would like to give a talk to the group, contact us to make arrangements. (Our addresses are below.)

How Do I obtain bionet.info-theory BY EMAIL?

If you have access to USENET news YOU DO NOT NEED AN E-MAIL SUBSCRIPTION!! We strongly encourage all interested users to explore getting USENET news at your site. It's MUCH easier on you than an e-mail subscription! Please consult your systems manager or contact biosci-help@net.bio.net for assistance if needed.

The BIOSCI (email) name for the forum is BIO-INFO.

Depending on where you are, you have to do different things to subscribe or be removed from the email subscription list:

SUBSCRIBING / UNSUBSCRIBING

North or South America or Pacific Rim:

Using the computer account in which you want to receive mail messages, please send an email message to the e-mail server at biosci-server@net.bio.net
Leave the Subject: line blank. In the body of the message include the line

subscribe bio-info

to add yourself to the mailing list or

unsubscribe bio-info

to cancel an existing subscription. If you need personal subscription assistance, please contact biosci-help@net.bio.net

Europe, Africa, and Central Asia:

Send a email message to the person at biosci@daresbury.ac.uk requesting a subscription or removal from the BIO-INFO forum.

SENDING OUT POSTINGS

Thereafter, address email messages for this forum to one of:

North or South America or Pacific Rim:
bio-info@net.bio.net

Europe, Africa, and Central Asia:
bio-info@daresbury.ac.uk

You can post to either of the above address if you want. We only request that you sign up at your local node in order to optimize the use of the network resources for message distribution.

Do not send subscription requests to any of these addresses, or you will have sent it to everybody on the planet (to your great embarrassment, and we will drub you with food cake)! Let me say that again: please do not post requests for subscription or being removed from the list to the list itself, that takes up bandwidth all over the world!

If you have problems, contact the subscription site manager who you signed up with. If your problem is not resolved, please contact biosci-help@net.bio.net

DO NOT CONTACT TOM SCHNEIDER FOR SUBSCRIPTIONS OR UNSUBSCRIBING!

This is so complicated! It would be a lot easier for you to use a news reader!

Where Did I Get This FAQ File Originally?

- The source **hypertext version** is available from http://www.ccrnp.ncifcrf.gov/~toms/bionet.info-theory.faq.html
- The **flatfile** version of this FAQ is stored in http://www.ccrnp.ncifcrf.gov/~toms/bionet.info-theory.faq.txt
- A **PostScript version** is available from http://www.ccrnp.ncifcrf.gov/~toms/bionet.info-theory.faq.ps
- A **PDF version** is available from http://www.ccrnp.ncifcrf.gov/~toms/bionet.info-theory.faq.pdf

If you are curious, all of these alternative forms are created automatically with a csh script called mkbitcs using features of netscape and the wonderful ImageMagick.

This file is posted monthly (if I remember to do it!) on news.answers and bionet.info-theory.

Please send questions and comments to: Tom Schneider (toms@ncifcrf.gov).

What is the IP number of the FAQ archive?

For ftp.ncifcrf.gov you can use "129.43.1.11", however, I am keeping everything on the web now.

Where Are the Bionet Archives?

The hypertext archives for this newsgroup are at:
http://www.bio.net/hypermail/BIOLOGICAL-INFORMATION-THEORY/

The entire collection of BIOSCI/bionet messages from inception are available via the biosci.src WAIS source at net.bio.net. Contact biosci-help@net.bio.net for further help with accessing this WAIS source.

Are There Other Archives?

- BIOSCI Archive of Monthly Postings. ftp://net.bio.net/pub/BIOSCI/BIOLOGICAL-INFORMATION-THEORY. This archive contains postings from each month as a single document. Files are in mailbox format, with names of the form YYMM (YY=last 2 digits of the year, MM=cardinal number of the month, zero padded). The current months postings are in the file 'current'. Contact biosci-help@net.bio.net for further help with or comments on the archives. For the record, the IP number for net.bio.net is [134.172.2.69].

- These are the BIOSCI raw postings, just numbered: ftp://net.bio.net/pub/BIOSCI/bionet/info-theory

- Archive of Postings at IUBO: ftp://ftp.bio.indiana.edu/usenet/bionet/info-theory/. This archive contains individual postings. Older postings are collected by the month as a single document. There is an index for each month.
- Archive of Life Related Newsgroups http://www.krl.caltech.edu/~brown/alife/news/This is an incredibly nicely organized HTML archive of links maintained by Titus Brown at Caltech (brown@krl.caltech.edu). This archive contains individual postings. Check it out!!
- current newsgroup articles on your own computer: bionet.info-theory
- The BIOSCI home page carries all bionet news groups: http://www.bio.net/

I Posted But Nothing Happened?!?

Michael Harman (rmharman@jhu.edu)

| I attempted to post a question ... about a
| month and a half ago, but never saw any response.

Go to the bionet archives

http://www.bio.net/hypermail/BIOLOGICAL-INFORMATION-THEORY/

and search for your posting. If your posting does not appear there within a day it may mean that your posting never made it out of your system. Try again to see if it was a transient failure. If that fails, talk to your systems admin. If your systems administrator is stumped, contact Dave Kristofferson at biosci-help@net.bio.net for further help. You could also check by posting on misc.test (it's fun, I promise! :-).

What is an Appropriate Posting?

Name calling and libelous statements are not acceptable on this news group. It's best to learn about net etiquette (netiquette) before you post anything.

On the other hand, polite, carefully worded, even aggressive **scientific criticism** that specifically addresses issues **is encouraged**. If you critique someone's work, be willing to defend your statements, and be willing to admit publically when you are wrong. When ad hominem postings appear, we will *quickly* conclude that you are intent on abusing us and will take appropriate, but legal, actions against you.

To maintain a high professional level of discussion, we encourage all participants to identify themselves. You do **not** need any degrees or professional affiliation to join the conversation, and you should not hesitate to post if you feel you have something worthwhile to contribute.

However, if you want to avoid looking naive, some knowledge about basic molecular biology and information theory also helps (see the references), but we don't expect you to be an expert on everything. Also, to make a good impression on others, trim any text you copy from previous postings, run your text through a spell checker, and use proper English.

What Can I Do About Inappropriate Postings?

The short form of this news group's name, bio-info, can be confusing to some people inexperienced in network communications or with little knowledge of the discipline (if there is any :-) of biological information theory. It can and has been mistaken as a news group for general biological information. Our readers should be aware that when such postings come to our attention, the discussion leaders do attempt to inform, privately, the people who make these inappropriate postings of the error of their ways and suggest alternative or more appropriate venues.

Subjecting the writers of inappropriate posting to public excoriation is not a good policy because it may be an inadvertent mistake and follow-up postings will only add to the irritation of our regular readers. When others publicly reply to such posts in this news group, although they may think they are being polite to the original poster, they are still annoying our regular readers. We suggest that a better policy for readers who do wish to reply to inappropriate posts is to do so privately or to an appropriate news group.

If you have nothing better to do with your time and feel you must reply to an inappropriate posting, either because you think it might be a sincere though misguided request for information, or because you want to express your opinions on the poster's ancestry, cool your jets one minute and carefully consider the poster's address. Look in the mail header for the "From:" line, the "Reply-to:" line, the "Message-id:" line, and the "Posting-Host:" line. If the "From:" or "Reply-to:" lines contain obviously forged information, like

From: Anonymous@net.bio.net (Unknown)
Reply-to: No.one.@net.bio.net

or if the address looks legitimate but contains inconsistent node addresses like

From: ReadMe@ReadMe.net
Message-id: <4upgib$af8@dfw-ixnews5.ix.netcom.com>

(the part after the "@" in these two lines is not consistent), do not waste your time. The poster will never read your reply. The posting is either a "spam" or an attempt to sabotage the system whose address has been forged.

More importantly, do not waste other scientists' time and money (yes, some people do pay for the e-mail they receive) by replying to an inappropriate posting through the bulletin board. No one else will be interested in seeing your inappropriate reply to an inappropriate posting. They may, however, note for future reference your lack of courtesy and good judgement.

For information about how to deal with intransigent cases, see:
http://math-www.uni-paderborn.de/~axel/blacklist.html

For dealing with Make Money Fast schemes, see:
http://www.ccrnp.ncifcrf.gov/~toms/mmf.html

Another anti-spam resource is at http://www.canismajor.demon.co.uk/antispam/antispam.htm

---

Should I send private email to someone to respond to a posting or to ask a question?

It's fine to email someone a question or comment about one of their postings, but remember that you will then be holding a private conversation with only that person and the rest of us will miss out on your thoughts and won't be able to help you. Of course, private email is appropriate if you are thinking of forming a collaboration with someone and don't want the ideas to be public, or if you have a technical question about the news group. Also, please don't post *and* send email to someone unless you have a good reason to think they will miss the posting.

In other words, please don't email to Tom Schneider general comments that could be public.

What is the official word on copyright of this FAQ?

This FAQ fits the description in the U. S. Copyright Act of a "United States Government work". It was written as a part of my official duties as Government employee. This means it cannot be copyrighted. The article is freely available without a copyright notice, and there are no restrictions on its use, now or subsequently. I retain no rights in the FAQ.

Thomas D. Schneider

Who Takes Care of This Group?

Steve Garavelli
Box 3783, Georgetown Station
Washington, DC 20007
202-625-1907
jsgaravelli@earthlink.net
http://home.earthlink.net/~jsgaravelli/

Tom Schneider
National Cancer Institute
Laboratory of Experimental and Computational Biology
Frederick, Maryland 21702-1201
toms@ncifcrf.gov
http://www.ccrnp.ncifcrf.gov/~toms/

John L. Spouge
National Center for Biotechnology Information
National Library of Medicine
Bethesda, MD 20894
spouge@ncbi.nlm.nih.gov
http://www.ncbi.nlm.nih.gov/CBBresearch/Spouge/

Please email comments and suggestions on this faq sheet to Tom.

John Garavelli (who also answers to "Steve" if you want to avoid confusion) often organizes dinner speakers.

John Spouge often arranges dinner locations.

What Kind of Questions Are Appropriate For Discussion?

This faq sheet answers simple questions about this group. The BIG questions should be discussed on the net, where we can all haggle over them. Here are a few for starters:

- What is the role of theory in biology today?
- What should be the role of biological theory?
- What is information? How should it be defined?
- What bothers you when you read the two papers on the theory of molecular machines? (It is only from the things that bother us that we can make progress in understanding.) (See references below.)
- What are flaws in the theory of molecular machines?
- How is ATP used to drive molecular machines?
- All communication systems are associated with living things, so is it true that information theory is really a theory about living things? Was Shannon really a great biologist?
- What does Maxwell's Demon have to do with all of this?
- What are the limits of computers?
- What are the limits of nanotechnology?
- Can we build molecular machines and molecular computers and how would they work?

When and Where are Meetings?

Meetings are announced in the bionet.info-theory news group. As of 1997 September 15, meetings and talks are announced at the Biological Information Theory and Chowder Society web page. If you know of are going to give a relevant talk, please submit information to Tom Schneider.

What is Information Theory?

Information theory is a branch of mathematics concerned with the process of making choices. Although it has a rich history going back centuries, it was the work of Claude Shannon, published in 1948 and later, that started the field. The theory is powerful and has resulted in great achievements. The beautiful sound we enjoy from compact disks (CD's) became possible only because of Shannon's work. The bionet.info-theory news group was formed to discuss the many applications of information theory to biology. (It is not a general information news group as some might be mislead to think.) It is worth at least some of your time to see why we are so excited about this application, as it could turn your research around by sharpening your experimental approaches.

Is There a Quick Introduction to Information Theory Somewhere?

See the primer on information theory:

ftp://ftp.ncifcrf.gov/pub/delila/primer.ps
or
http://www.ccrnp.ncifcrf.gov/~toms/paper/primer

I'm Confused: How Could Information Equal Entropy?

If someone says that information = uncertainty = entropy, then they are confused, or something was not stated that should have been. Those equalities lead to a contradiction, since entropy of a system increases as the system becomes more disordered. So information corresponds to disorder according to this confusion.

If you always take information to be a decrease in uncertainty at the receiver and you will get straightened out:

R = Hbefore - Hafter.

where H is the Shannon uncertainty:

H = - sum (from i = 1 to number of symbols) Pi log2 Pi (bits per symbol)

and Pi is the probability of the ith symbol. If you don't understand this, please refer to "Is There a Quick Introduction to Information Theory Somewhere?".

Imagine that we are in communication and that we have agreed on an alphabet. Before I send you a bunch of characters, you are uncertain (Hbefore) as to what I'm about to send. After you receive a character, your uncertainty goes down (to Hafter). Hafter is never zero because of noise in the communication system. Your decrease in uncertainty is the information (R) that you gain.

Since Hbefore and Hafter are state functions, this makes R a function of state. It allows you to lose information (it's called forgetting). You can put information into a computer and then remove it in a cycle.

Many of the statements in the early literature assumed a noiseless channel, so the uncertainty after receipt is zero (Hafter=0). This leads to the SPECIAL CASE where R = Hbefore. But Hbefore is NOT "the uncertainty", it is the uncertainty of the receiver BEFORE RECEIVING THE MESSAGE.

A way to see this is to work out the information in a bunch of DNA binding sites.

Definition of "binding": many proteins stick to certain special spots on DNA to control genes by turning them on or off. The only thing that distinguishes one spot from another spot is the pattern of letters (nucleotide bases) there. How much information is required to define this pattern?

Here is an aligned listing of the binding sites for the cI and cro proteins of the bacteriophage (i.e., virus) named lambda:

```
alist 5.66 aligned listing of:
* 96/10/08 19:47:44, 96/10/08 19:31:56, lambda cI/cro sites
piece names from:
* 96/10/08 19:47:44, 96/10/08 19:31:56, lambda cI/cro sites
The alignment is by delila instructions
The book is from:   -101 to 100
This alist list is from: -15 to 15

                       ------                   ++++++
                       111111--------- +++++++++111111
                       5432109876543210123456789012345
                       ...............................
OL1 J02459  35599 +  1 tgctcagtatcaccgccagtggtatttatgt
    J02459  35599 -  2 acataaataccactggcggtgatactgagca
OL2 J02459  35623 +  3 tttatgtcaacaccgccagagataatttatc
    J02459  35623 -  4 gataaattatctctggcggtgttgacataaa
OL3 J02459  35643 +  5 gataatttatcaccgcagatggttatctgta
    J02459  35643 -  6 tacagataaccatctgcggtgataaattatc
OR3 J02459  37959 +  7 ttaaatctatcaccgcaagggataaatatct
    J02459  37959 -  8 agatatttatcccttgcggtgatagatttaa
OR2 J02459  37982 +  9 aaatatctaacaccgtgcgtgttgactattt
    J02459  37982 - 10 aaatagtcaacacgcacggtgttagatattt
OR1 J02459  38006 + 11 actattttacctctggcggtgataatggttg
    J02459  38006 - 12 caaccattatcaccgccagaggtaaaatagt
                                             ^
```

Each horizontal line represents a DNA sequence, starting with the 5’ end on the left, and proceeding to the 3’ end on the right. The first sequence begins with: 5’ tgctcag ... and ends with ... tttatgt 3’. Each of these twelve sequences is recognized by the lambda repressor protein (called cI) and also by the lambda cro protein.

What makes these sequences special so that these proteins like to stick to them? Clearly there must be a pattern of some kind.

Read the numbers on the top vertically. This is called a "numbar". Notice that position +7 always has a T (marked with the ^). That is, according to this rather limited data set, one or both of the proteins that bind here always require a T at that spot. Since the frequency of T is 1 and the frequencies of other bases there are 0, H(+7) = 0 bits. But that makes no sense whatsoever! This is a position where the protein requires information to be there.

That is, what is really happening is that the protein has two states. In the BEFORE state, it is somewhere on the DNA, and is able to probe all 4 possible bases. Thus the uncertainty before binding is Hbefore = log2(4) = 2 bits. In the AFTER state, the protein has bound and the uncertainty is lower: Hafter(+7) = 0 bits. The information content, or sequence conservation, of the position is Rsequence(+7) = Hbefore - Hafter = 2 bits. That is a sensible answer. Notice that this gives Rsequence close to zero outside the sites.

If you have uncertainty and information and entropy confused, I don't think you would be able to work through this problem. For one thing, one would get high information OUTSIDE the sites. Some people have published graphs like this.

A nice way to display binding site data so you can see them and grasp their meaning rapidly is by the sequence logo method. The sequence logo for the example above is at http://www.ccrnp.ncifcrf.gov/~toms/gallery/hawaii.fig1.gif. More information on sequence logos is in the section What are Sequence Logos?

More information about the theory of BEFORE and AFTER states is given in the papers http://www.ccrnp.ncifcrf.gov/~toms/paper/nano2, http://www.ccrnp.ncifcrf.gov/~toms/paper/ccmm and http://www.ccrnp.ncifcrf.gov/~toms/paper/edmm.

How Can I Learn More About Information Theory and Biology? References

REFERENCES - General

There are a huge number of papers related to this topic, just about everything in molecular biology, lots of chemistry, physics, electronics, evolutionary theory, thermodynamics, statistical mechanics and the kitchen sink ... References are given in BiBTeX format, the bibliography program associated with LaTeX, the powerful and portable typesetting program.

By arrangement, books that have prices listed can be ordered over Internet from:

Reiter's Scientific & Professional Books
2021 K Street, NW
Washington, DC 20006
1-800-537-4314
1-202-223-3327
1-202-296-9103 FAX
EMAIL:
books@reiters.com
WWW:
http://reiters.com/

Shipping and handling charges are: in the DC metropolitan area $4.00 for one item, $0.50 for each additional item, outside the area $4.50 for one item, $0.50 for each additional item.

The prices are current as of October 1994; because publishers are constantly changing their prices, they should be considered estimates rather than guaranteed prices. To open an account you must first either phone or FAX them and provide a credit card number. Book orders can be then placed at any time over the Internet. **DO NOT SEND CREDIT CARD NUMBERS OVER THE INTERNET!**

Reiter's carries all of the books on this list except "Information Theory: Saving Bits", and that one can be special ordered. If enough interest in this book is generated by the FAQ, it will be added as regular stock. (It can also be ordered directly from the company using the information given.)

Gonick's Wonderful books (Don't be shy! They are worth the money!!):

```
@book{Gonick.computers,
author = "L. Gonick",
title = "The Cartoon Guide to Computers",
edition = "second",
publisher = "HarperCollins",
address = "New York, NY",
isbn = "0-06-273097-5",
price = "price as of 1994 October 31: \$11.00",
year = "1991"}
```

```
@book{Gonick.genetics,
author = "L. Gonick",
title = "The Cartoon Guide to Genetics",
edition = "updated",
publisher = "Barnes \& Nobel",
address = "New York, NY",
isbn = "0-06-273099-1",
price = "price as of 1994 October 31: \$12.00",
year = "1991"}
```

```
@book{Gonick.physics,
author = "L. Gonick
and A. Huffman",
title = "The Cartoon Guide to Physics",
publisher = "HarperPerennial",
address = "New York, NY",
isbn = "0-06-273100-9",
price = "price as of 1994 October 31: \$12.00",
year = "1990"}
```

A good starting point if you don't know much molecular biology: (Two volumes)

```
@book{Watson1987,
author = "J. D. Watson
and N. H. Hopkins
and J. W. Roberts
and J. A. Steitz
```

and A. M. Weiner",
title = "Molecular Biology of the Gene",
edition = "fourth",
publisher = "The Benjamin/Cummings Publishing Co., Inc.",
address = "Menlo Park, California",
isbn = "0-8053-9614-4",
price = "price as of 1994 October 31: \$59.95",
year = "1987"}

This book describes LaTex and BiBTeX:

@book{Lamport1994,
author = "L. Lamport",
title = "\LaTeX: A Document Preparation System,
User's Guide \& Reference Manual",
edition = "second",
publisher = "Addison-Wesley Publishing Company",
address = "Reading, Massachusetts",
isbn = "0-201-52983-1",
price = "price as of 1994 October 31: \$32.95",
year = "1994"}

REFERENCES - Information Theory

- **Basic References**
  - John Pierce was at Bell Labs while Shannon dreamed up information theory. He saw the development from the inside, and wrote it up in "An Introduction to Information Theory: Symbols, Signals and Noise". Although it is not highly mathematical, this book is still the best one to start with because it gives one a feeling for the scope and implications of the theory, without dumping on the math, yet without leaving out important topics that later generations of popular writers skipped.

    @book{Pierce1980,
    author = "J. R. Pierce",
    title = "An Introduction to Information Theory:
    Symbols, Signals and Noise",
    edition = "second",
    publisher = "Dover Publications, Inc.",
    address = "New York",
    isbn = "0-486-24061-4",
    comment = "
    original copyright 1961
    Ordering information: Pierce1980 is currently available by mail from:
    Dover Publications, Inc.
    31 East 2nd street
    Mineola, New York 11501

order:
Pierce, An Introduction to Information Theory: Symbols, Signals and Noise
code number: 24061-4
$7.95 + charges. Payment in full, no telephone or credit card orders.
Postage and Handling charges are:
Bookrate: $3 (US only)
UPS: $4.50 (US only, not Alaska or Hawaii or PO boxes)
Foreign orders: add 20% of total (minimum $2.50)
Sales Tax (Ny residents only)
Foreign Orders Note: Remittances must be sent by international money order or in U.S. funds via Federal Wire System to Chemical Bank, N. Y. ABA #021000128. Mark all remittances 'For the account of Dover Publications, Inc. #001 053 272'. This information is from the Dover Math and Science Catalogue 9/92", price = "price as of 1994 October 31: \$8.95", year = "1980"}

- Christopher Hillman (hillman@math.washington.edu) suggests that Cover and Thomas' book is a better starting point, but that's because he is a mathematician. People who have seen both could post their opinions.

  @book{Cover.Thomas1991,
  author = "Thomas M. Cover
  and Joy A. Thomas",
  title = "Elements of Information Theory",
  publisher = "John Wiley \& Sons, Inc.",
  address = "N. Y.",
  isbn = "0-471-06259-6",
  year = "1991"}

- A good introduction to the mathematics, written for high school students:

  @book{Sacco1988,
  author = "W. Sacco
  and W. Copes
  and C. Sloyer
  and R. Stark",
  title = "Information Theory: Saving Bits",
  publisher = "Janson Publications, Inc.",
  comment = "original address was Providence, Rhode Island",
  address = "Dedham, MA",
  isbn = "0-939765-25-X",
  phone = "(800) 322-6284",
  price = "price as of 1994 October 31: \$11.95",
  year = "1988"}

- **Important originals:**

  - @article{Shannon1948,
    author = "C. E. Shannon",

title = "A Mathematical Theory of Communication",
year = "1948",
journal = "Bell System Tech. J.",
volume = "27",
pages = "379-423, 623-656"}
On line version at Bell Labs

- @book{ShannonWeaver1949,
  author = "C. E. Shannon
  and W. Weaver",
  title = "The Mathematical Theory of Communication",
  publisher = "University of Illinois Press",
  address = "Urbana",
  isbn = "0-252-72548-4",
  price = "price as of 1994 October 31: \$9.95",
  year = "1949"}

- @article{Shannon1949,
  author = "C. E. Shannon",
  title = "Communication in the Presence of Noise",
  year = "1949",
  journal = "Proc. IRE",
  volume = "37",
  pages = "10-21"}

- For the committed: The Complete Works!

  @book{Sloane.Wyner1993,
  author = "N. J. A. Sloane and A. D. Wyner",
  title = "Claude Elwood Shannon: Collected Papers",
  publisher = "IEEE Press",
  address = "Piscataway, NJ",
  isbn = "0-7803-0434-9",
  comment = "IEEE Order Number: PC0331-9
  ll To order directly by charge card (eg Visa works) you can call (908)-981-0060
  $69.95 + $5 handling charge
  delivery in about 2 weeks",
  price = "price as of 1994 October 31: \$69.95",
  comment = "this was previously called Shannon1993",
  year = "1993"}

- **Other basic references**

  - How locks work and other cool stuff:

    @book{Macaulay1988,
    author = "D. Macaulay",
    title = "The Way Things Work",

publisher = "Houghton Mifflin Company",
address = "Boston",
isbn = "0-395-42857-2",
price = "price as of 1994 October 31: \$29.95",
comment = "This book is also available on Windows-Compatible CD-ROM
cdrom isbn = 1-56458-901-3 Price as of 1994 October 31: \$99.95",
year = "1988"}

- Leff1990 gives a review of the Maxwell's Demon problem.
  See also Schneider.edmm, listed below.

@book{Leff1990,
author = "H. S. Leff and A. F. Rex",
title = "Maxwell's Demon: Entropy, Information, Computing",
publisher = "Princeton University Press",
address = "Princeton, N. J.",
phone = "1(800) 777-4726",
isbn.hard = "0-691-08726-1 (hard cover)",
price.hard = "price as of 1994 October 31: \$80.00",
isbn.paper = "0-691-08727-X (paperback)",
price.paper = "price as of 1994 October 31: \$26.95",
year = "1990"}

REFERENCES - Jaynes

@article{JaynesI,
author = "Edwin T. Jaynes",
title = "Information Theory and Statistical Mechanics",
year = 1957,
journal = "Physical Review",
volume = "106",
pages = "620-630"}

@article{JaynesII,
author = "Edwin T. Jaynes",
title = "Information Theory and Statistical Mechanics. {II}",
year = 1957,
journal = "Physical Review",
volume = "108",
pages = "171-190"}

A version of Jaynes' new book "PROBABILITY THEORY -- THE LOGIC OF SCIENCE" is available on the net. See:

ftp://bayes.wustl.edu/Jaynes.book/
Larry Bretthorst (larry@bayes.wustl.edu)

http://omega.albany.edu:8008/JaynesBook.html
Carlos Rodriguez (carlos@math.albany.edu)

Tom Schneider's pointers to these places:
http://www.ccrnp.ncifcrf.gov/~toms/jaynes.html

Note: The book is being written now and new versions come out every once in a while. One of these locations may be more up to date than the other.

REFERENCES - Schneider

To see online papers, go to http://www.ccrnp.ncifcrf.gov/~toms/paper

@article{Schneider1986,
author = "T. D. Schneider
and G. D. Stormo
and L. Gold
and A. Ehrenfeucht",
title = "Information content of binding sites on nucleotide sequences",
journal = "J. Mol. Biol.",
volume = "188",
pages = "415-431",
year = "1986"}

@inproceedings{Schneider1988,
author = "T. D. Schneider",
editor = "G. J. Erickson and C. R. Smith",
title = "Information and entropy of patterns in genetic switches",
booktitle = "Maximum-Entropy and Bayesian Methods in Science and Engineering",
volume = "2",
pages = "147-154",
publisher = "Kluwer Academic Publishers",
address = "Dordrecht, The Netherlands",
year = "1988"}

@article{Schneider1989,
author = "T. D. Schneider
and G. D. Stormo",
title = "Excess Information at Bacteriophage {T7} Genomic Promoters
Detected by a Random Cloning Technique",
year = "1989",
journal = "Nucl. Acids Res.",
volume = "17",
pages = "659-674"}

@article{Schneider.Stephens.Logo,
author = "T. D. Schneider
and R. M. Stephens",
title = "Sequence Logos: A New Way to Display Consensus Sequences",
journal = "Nucl. Acids Res.",
volume = "18",
pages = "6097-6100",
year = "1990"}

@article{Schneider.ccmm,
author = "T. D. Schneider",
title = "Theory of Molecular Machines.
{I. Channel} Capacity of Molecular Machines",
journal = "J. Theor. Biol.",
volume = "148",
number = "1",
pages = "83-123",
note = "{(Note: The figures were printed out of order!
Fig. 1 is on p. 97.)}",
year = 1991}

@article{Schneider.edmm,
author = "T. D. Schneider",
title = "Theory of Molecular Machines.
{II. Energy} Dissipation from Molecular Machines",
journal = "J. Theor. Biol.",
volume = "148",
number = "1",
pages = "125-137",
year = 1991}

@article{Herman.Schneider1992,
author = "N. D. Herman
and T. D. Schneider",
title = "High Information Conservation Implies that at Least Three Proteins Bind Independently to {F}
Plasmid {{\em incD\/}} Repeats",
journal = "J. Bact.",
volume = "174",
pages = "3558-3560",
year = "1992"}

@article{Stephens.Schneider.Splice,
author = "R. M. Stephens
and T. D. Schneider",
title = "Features of spliceosome evolution and function
inferred from an analysis of the information at human splice sites",
journal = "J. Mol. Biol.",
volume = "228",

pages = "1124-1136",
year = "1992"}

@article{Papp.helixrepa,
author = "P. P. Papp
and D. K. Chattoraj
and T. D. Schneider",
title = "Information Analysis of Sequences that Bind the Replication Initiator {RepA}",
journal = "J. Mol. Biol.",
comment = "Cover of 233, number 2!",
volume = "233",
pages = "219-230",
year = "1993"}

@article{Schneider.nano2,
author = "T. D. Schneider",
title = "Sequence Logos, Machine/Channel Capacity,
{Maxwell}'s Demon, and Molecular Computers: a Review of the Theory of Molecular Machines",
journal = "Nanotechnology",
volume = "5",
number = "1",
pages = "1-18",
year = "1994"}
ftp://ftp.ncifcrf.gov/pub/delila/nano2.ps

REFERENCES - Yockey

@book{Yockey1958a,
editor = "Hubert P. Yockey and Robert P. Platzman and Henry Quastler",
title = "Symposium on Information Theory in Biology",
booktitle = "Symposium on Information Theory in Biology",
publisher = "Pergamon Press",
address = "New York, London",
comment = "out of print",
year = "1958"}

@article{Yockey1981,
author = "Hubert P. Yockey",
year = 1981,
title = "Self-organization Origin of Life Scenarios and Information Theory",
journal = "J. Theor. Biol.",
volume = "91",
pages = "13-31"}

@book{Yockey1992,
author = "H. P. Yockey",

title = "Information Theory in Molecular Biology",
publisher = "Cambridge University Press",
address = "Cambridge",
isbn = "0-521-35005-0",
comment = "40 West 20th Street,
New York, N. Y. 10011-4211,
order number 350050",
phone = "1-800-827-7423",
price = "price as of 1994 October 31: \$74.95",
year = "1992"}

Following is Hubert Yockey's reference list:

- Yockey, Hubert P. Information Theory and Molecular Biology, Cambridge UK: Cambridge University Press (1992)

- When is random random? Nature 344 (1990) p823, Hubert P. Yockey

- Yockey, Hubert P. (1981). Self-organization origin of life scenarios and information theory. Journal of Theoretical Biology, 91, 13-31.

- Yockey, Hubert P. (1979). Do overlapping genes violate molecular biology and the theory of evolution? Journal of Theoretical Biology, 80, 21-26.

- Yockey, Hubert P. (1978). Can the Central Dogma be derived from information theory? Journal of Theoretical Biology, 74, 149-152.

- Yockey, Hubert P. (1977a). A prescription which predicts functionally equivalent residues at given sites in protein sequences. 67, 337-343.

- Yockey, Hubert P. (1977b). On the information content of cytochrome c. Journal of Theoretical Biology, 67, 345-376.

- Yockey, Hubert P. (1977c). A calculation of the probability of spontaneous biogenesis by information theory. Journal of Theoretical Biology, 67, 377-398.

- Yockey, Hubert P (1974). An application of information theory to the Central Dogma and the sequence hypothesis. Journal of Theoretical Biology,.46, 369-406.

- Yockey, Hubert P. (1960) The Use of Information Theory in Aging and Radiation Damage In The Biology of Aging American Institute of Biological Sciences Symposium No. 6 (160) pp338-347.

- Yockey, Hubert P., Platzman, Robert P. & Quastler, Henry, eds. (1958a). Symposium on Information Theory in Biology, New York, London: Pergamon Press.

- Yockey, Hubert P. (1958b). A study of aging, thermal killing and radiation damage by information theory. In Symposium on Information Theory in Biology. eds. Hubert P. Yockey, Robert Platzman & Henry Quastler, pp297-316. New York,London: Pergamon Press.

- Yockey, Hubert P. (1956). An application of information theory to the physics of tissue damage. Radiation.Research, 5, 146-155.
- Information in bits and bytes; Reply to Lifson's Review of "Information Theory and Molecular" Biology BioEssays v17 p85-88 (1995)
- Comments on "Let there be life; Thermodynamic Reflections on Biogenesis and Evolution by Avshalom C. Elitzur Journal of Theoretical Biology in press (1995).

REFERENCES - Adleman and papers related to molecular computation

Tom Schneider has a list of molecular computation resources.

A longer and more complete list of references is maintained by J.H.M.Dassen (jdassen@wi.leidenuniv.nl) in A biblography on Molecular Computation and Splicing Systems (http://www.wi.LeidenUniv.nl/~jdassen/dna.bib)There are also hyperlinks to most of the (90+) papers (http://www.wi.LeidenUniv.nl/~jdassen/dna.html)

@article{Adleman1994,
author = "Leonard M. Adleman",
title = "Molecular computation of solutions to combinatorial problems",
journal = "Science",
volume = "266",
pages = "1021-1024",
date = "November 11",
year = 1994}

@article{Baum1995,
author = "Eric B. Baum",
title = "Building an associative memory vastly larger that the brain",
journal = "Science",
volume = "268",
pages = "583-585",
date = "April 28",
year = 1995}

@article{Lipton1995,
author = "Richard J. Lipton",
title = "DNA solution of hard computational problems",
journal = "Science",
volume = "268",
pages = "542-545",
date = "April 28",
year = 1995}

@manuscript{Adleman1995,

author = "Leonard M. Adleman",
title = "On constructing a molecular computer",
note = "Available by anonymous ftp:
/pub/csinfo/papers/adleman/molecular_computer.ps on usc.edu",
year = 1995}

Other available manuscripts:

1. Dick Lipton of Princeton
Speeding up computations via molecular biology. Draft. Dec. 9, 1994.
ftp://ftp.cs.princeton.edu/pub/people/rjl/bio.ps

2. Dan Boneh of Princeton has several manuscripts available at:
Breaking DES Using a Molecular Computer.
Authors: D. Boneh, C. Dunworth, R. Lipton
This paper contains the talk from the workshop.
http://www.cs.princeton.edu/~dabo/biocomp.html

On the Computational Power of DNA.
Authors: D. Boneh, C. Dunworth, R. Lipton, J. Sgall
This is a new paper which contains several results:
a. Shows how to solve the circuit satisfaction problem.
b. Shows how to solve optimization problems such as MAX-Clique without going through decision problems.
c. Shows how to evaluate predicates in the polynomial hirarchy.

Making DNA Computers Error Resistant.
Authors: D. Boneh, R. Lipton
This paper shows how to transform volume reducing DNA algorithms into algorithm that are resistant to errors.

REFERENCES - Gad Yagil and papers related to Algorithmic Information Theory (AIT) or Algorithmic Complexity

An alternative way to analyze biosystems is by the Algorithmic Information Theory (AIT) or Algorithmic Complexity (AC) approach, first formulated by Kolmogoroff, Solomonoff and Chaitin in the 1960's. According to this approach, the information in a string of symbols is equal to the length of the shortest program caparisons of reproducing the string. This concept has been reformulated to tackle real molecular and biosystems ("Structural Complexity") and applied to a range of biosystems by G. Yagil. The more recent publications, which include references to the work of Kolmogoroff and of Chaitin, can be found at:

http://www.weizmann.ac.il/~lcyagil
also at http://interjournal.org, Manuscript Number 135. (Do a search for the manuscript number.)

The book of Cover and Thomas covers AC extensively. In particular, it shows that under certain conditions, AC can become equal to the Shannon information (or uncertainty) measure. In a series of papers, C.H. Bennett has proposed a concept of "logical depth", related to the time required by a universal machine to compute a sequence, as another measure of the information content of a string or sequence:

see: C.H. Bennett, "Logical Depth and Physical Complexity". In: "The Universal Turing Machine -A half century", Rolf Herken, Editor, Oxford University press, 1988.

Gad Yagil, Ph. D.
Dept. of Molecular Cell Biology
The Weizmann Institute of Science
Rehovot, Israel, 76100
Tel. 089-460-918 (home)
Fax 089-344-125
e-mail lcyagil@wiccmail.weizmann.ac.il.

REFERENCES - Entropy on the World Wide Web.

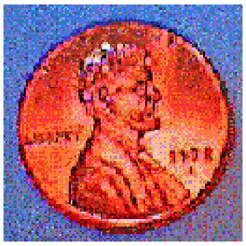
- Entropy on the World Wide Web. Chris no longer maintains the page, it is maintained by Roland Gunesch (Mathematics, Penn State). http://www.math.psu.edu/gunesch/entropy.html
- Chris Hillman's Home Page: http://www.math.washington.edu/~hillman/personal.html

---

Will Authors Send Me Papers?

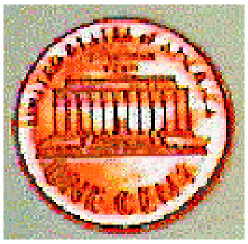
Tom Schneider will mail you copies of some of his papers. You can request them through the World Wide Web from http://www.ccrnp.ncifcrf.gov/~toms/papers.html or by sending your physical address to him at toms@ncifcrf.gov.

If you are willing to send out papers or have papers you would like listed here, please contact Tom Schneider.



Where Can I Get BIG Coins?

BIG coins are nice for explaining that a bit represents the choice between two equally likely possibilities.

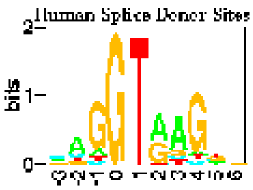


News Emporium, Inc. (703) 661-3550 sells large coins at Dulles International Airport.

Parks and History has big coins for sale. They will have a web site Bookshop soon. In the meantime, you could call (202) 755-0461 or (800) 990-7275. They accept VISA, MasterCard or American Express. Contact: Linda Depew their Mail Order & Wholesale Manager.

If you find other sources, please tell Tom Schneider.

What are Sequence Logos?



A sequence logo is a graphical method for showing patterns created by using information theory.

How Do I find Sequence Logos on the Web?

http://www.ccrnp.ncifcrf.gov/~toms/sequencelogo.html

Is There a Shell Script for Making Sequence Logos?

Yes, you will find the one Shmuel Pietrokovski wrote in the ftp archive ftp.ncifcrf.gov in pub/delila/logoaid. (Also available in bioinformatics.weizmann.ac.il/pub/software/logoaid.)

Is There a World Wide Web Page for Making Sequence Logos?

Yes, Steve Brenner has done it!

http://www.bio.cam.ac.uk/seqlogo/

Are There Other Organizations for Information Theory?

IEEE Information Theory Society

Acknowledgments

This FAQ is written and maintained by Tom Schneider. It was HTMLized by Susan Hogarth (sjhogart@unity.ncsu.edu) in February, 1997 but is NOT maintained by her. Please look at Who Takes Care of This Group if you have questions about this FAQ.