Probabilistic Graphical Models in Bioinformatics

Tutorial 4: Introduction to Information Theory, and JAGS

Fahimeh Palizban
Department of Bioinformatics, IBB, University of Tehran

26 Farvardin,1398
(15 April ,2019)

University of TEHRAN

Contents

- What is information theory?
- Entropy
- Conditional joint entropy
- Mutual information
- Kullback–Leibler divergence
- Gibbs sampling
- Markov Chain Monte Carlo
- Introduction to JAGS

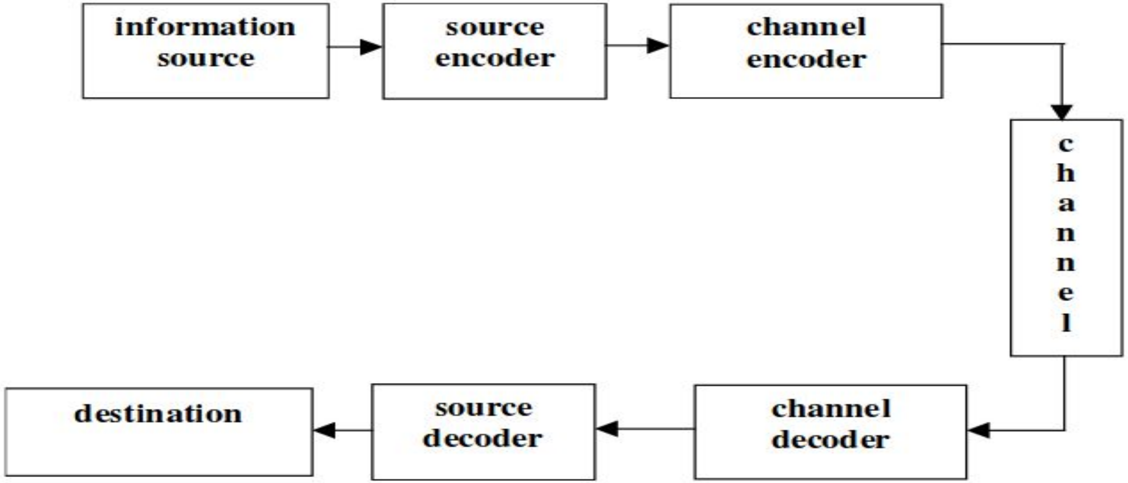
Information Theory

Definition of information theory

- **Information theory** studies the quantification, storage, and communication of information. It was originally proposed by Claude Shannon in 1948.

Information theory also provides methodologies to separate real information from noise and to determine the channel capacity required for optimal transmission conditioned on the transmission rate.

- Shannon’s information measures refer to entropy, conditional entropy, mutual information,and conditional mutual information.



Entropy

The average amount of information in observing the output of the source.

$$H(S) = \sum_i p_i \cdot I(s_i) = \sum_i p_i \cdot \log \frac{1}{p_i} = E_P\left[\log \frac{1}{p(s)}\right]$$

avg amt of info provided per symbol2.

avg amount of surprise when observing a symbol.

uncertainty an observer has before seeing the next symbol.

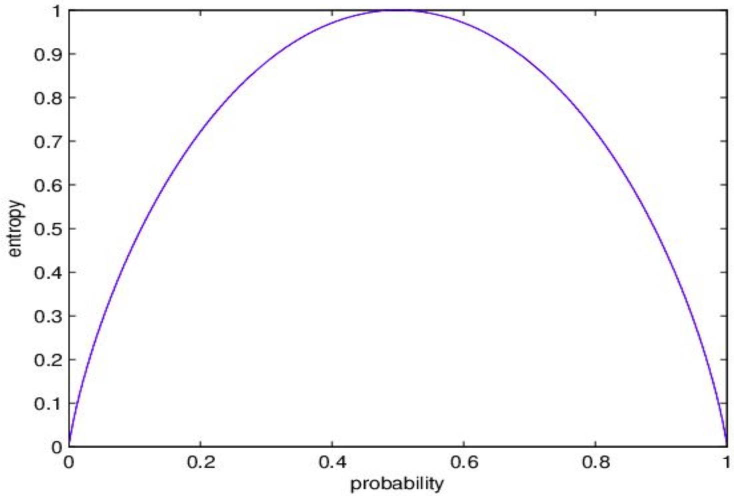
avg of bits needed to communicate each symbol

Binary entropy function

H(p), is defined as the entropy of a Bernoulli process with probability p of one of two values.

If $\Pr(X=1)=p$, then $\Pr(X=0)=1-p$ and the entropy of X (in shannons) is given by

$$\mathrm{H}(X)=\mathrm{H}_{\mathrm{b}}(p)=-p\log_2 p-(1-p)\log_2(1-p),$$



Joint Entropy

joint entropy is a measure of the uncertainty associated with a set of variables

$$H(X,Y) = -\sum_{S_X}\sum_{S_Y} p(x,y)\log p(x,y)$$

the general expression for the decomposition of joint entropy is :

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$

Conditional Entropy

conditional entropy (or equivocation) quantifies the amount of information needed to describe the outcome of a random variable given that the value of another random variable is known.

The conditional entropy of Y given X is defined as:

$$H(Y|X) = -E[\log p(Y|X)] = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x)$$

Note that :

$$\begin{aligned} H(Y|X) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) = -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x) \\ &= \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \end{aligned}$$

Mutual Information

the **mutual information** (**MI**) of two random variables is a measure of the mutual dependence between the two variables

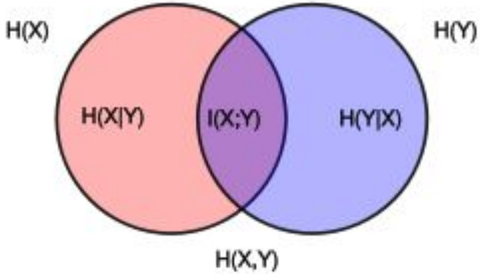
More specifically, it quantifies the "amount of information" (in units such as shannons, commonly called bits) obtained about one random variable through observing the other random variable.

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = E \log \frac{p(X,Y)}{p(X)p(Y)}.$$

Relation to conditional and joint entropy

Mutual information can be equivalently expressed as

$$\begin{aligned} \mathrm{I}(X;Y) &\equiv \mathrm{H}(X) - \mathrm{H}(X|Y) \\ &\equiv \mathrm{H}(Y) - \mathrm{H}(Y|X) \\ &\equiv \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y) \\ &\equiv \mathrm{H}(X,Y) - \mathrm{H}(X|Y) - \mathrm{H}(Y|X) \end{aligned}$$



Kullback–Leibler divergence

is a non-symmetric measure of the difference between two probability distributions p(x) andq(x)

The relative entropy or Kullback-Leibler divergence between two distributions,p(x) andq(x) :

$$D_{KL}(p(x)||q(x)) = \sum_{x \in X} p(x) \ln \frac{p(x)}{q(x)}$$

The KL divergence measures the expected number of extra bits required to code samples from p(x) when using a code based onq(x), rather than using a code based on p(x)

Example: Consider two different Bernoulli distributions.

Bayesian Inference: Gibbs Sampling

Gibbs Sampling

The main idea is to break the problem of sampling from the high-dimensional joint distribution into a series of samples from low-dimensional conditional distributions.

The dependence of the samples turns out to follow a Markov distribution, leading to the name Markov chain Monte Carlo (MCMC) which attempt to simulate direct draws from some complex distribution of interest.

Algorithm 1 Gibbs sampler

Initialize $x^{(0)} \sim q(x)$
for iteration $i = 1, 2, \ldots$ **do**
$x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)}, \ldots, X_D = x_D^{(i-1)})$
$x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)}, \ldots, X_D = x_D^{(i-1)})$
$\vdots$
$x_D^{(i)} \sim p(X_D = x_D | X_1 = x_1^{(i)}, X_2 = x_2^{(i)}, \ldots, X_D = x_{D-1}^{(i)})$
end for

Bayesian statistical inference

There are several state-of-the-art platforms for statistical modeling and high-performance statistical computation.

These tools can be used for statistical modeling, data analysis, and prediction in the social, biological, and physical sciences, engineering, and business.

Most commonly used tools:

Just another Gibbs sampler (JAGS)

WinBUGS

OpenBUGS

STAN

What is JAGS?

It is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation.

It uses a dialect of the BUGS language, similar but a little different to OpenBUGS and WinBUGS.

It is written in C++.

JAGS runs on Linux, Mac, and Windows through an interface with R called rjags.

JAGS Installation

The following sets out a basic installation process:

1. If necessary Download and install R and potentially a user interface to R like R Studio
2. Download and install JAGS as per operating system requirements.
3. Install additional R packages: e.g., in R install.packages("rjags") to interface with JAGS and coda to process MCMC output

Running a JAGS model from R

Running a JAGS model refers to generating samples from the posterior distribution of the model parameters.

This takes place in five steps:

1. Definition of the model

2. Compilation

3. Initialization

4. Adaptation and burn-in

5. Monitoring

References

http://iest2.ie.cuhk.edu.hk/~whyeung/post/draft7.pdf

http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/diuf.unifr.ch/tcs/courses/it04-05/script/information-theory.pdf

https://web.stanford.edu/~montanar/RESEARCH/BOOK/partA.pdf

https://gtas.unican.es/files/docencia/TICC/apuntes/tema1awp.pdf

nitro.biosci.arizona.edu/courses/EEB596/handouts/Gibbs.pdf

https://www4.stat.ncsu.edu/~reich/st740/Computing2.pdf

https://www.r-bloggers.com/getting-started-with-jags-rjags-and-bayesian-modelling/

University of TEHRAN

Thanks!