Lecture: Construction of Biological Networks

Martin Vingron

MPI for Molecular Genetics

June-July 2017

Reverse engineering of biological networks

What we are going to cover:

- Networks for modeling biological relationships
- Introduce/refresh some basic concepts of probability
- Regression and correlation
- Information theory
- Partial correlation and the inverse variance-covariance matrix
- Markov random-fields and Gaussian Graphical Models
- Bayes Networks

Literature

Bishop: Pattern Recognition and Machine Learning, Springer esp. Chapter 8 Graphical Models;
Other reading: Whittacker; Wermuth and Cox; Lauritzen, EOSL chapter 17; Koller-Friedman book;
`http://swoh.web.engr.illinois.edu/courses/IE598/`
`https://www.coursera.org/specializations/probabilistic-graphical-models`

Networks

In biology, the term „network“ describes a graph whose nodes correspond to biological entities (e.g., genes) and whose edges indicate a relationship between the entities, e.g., regulatory interactions among genes ("regulatory network"), or physical interactions between proteins ("protein-protein interaction", PPI, network).

Sometimes the graph comes with edge-weights, sometimes without.

Gene expression data

Gene expression data have frequently given rise to gene networks. Gene expression data denotes a collection of mRNA levels for many genes across a number of conditions (cell-types, tissues, treatments).

The idea is that the expression levels of the genes indicate connections between genes, when, e.g., two genes tend to be up-regulated in the same tissues and down-regulated the others.

The abstract idea behind this construction is that a node represents a random variable (the mRNA levels), for which we observe a sample, namely the result of the gene expression experiment.

Then the question, whether there exists a relationship between the variables, becomes: Are the distributions at the nodes stochastically independent?

What we are going to cover:

- Networks for modeling biological relationships
- **Introduce/refresh some basic concepts of probability**
- Regression and correlation
- Information theory
- Partial correlation and the inverse variance-covariance matrix
- Markov random-fields and Gaussian Graphical Models
- Bayes Networks

Stochastic independence

See Wikipedia article "Independence (probability theory)". Review the notions of "joint distribution" and "marginal distribution". In a nutshell, for events:
Conditional probability:

$$P(B|A) = \frac{P(AB)}{P(A)}$$

If A and B are **independent**, then A has no influence on B (and B has no influence on A). Therefore, conditioning on A (or B) makes no difference: $P(A) = P(A|B)$, and $P(B) = P(B|A)$.
Therefore, A and B are independent iff:

$$P(AB) = P(B)P(A|B) = P(B|A)P(A) = P(A)P(B)$$

or for densities:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

.

Product rule (aka Chain rule)

$$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

or

$$P(A, B, C) = P(A)P(B|A)P(C|A, B)$$

etc.

For two independent variables, the joint density is the product of their marginals. See:
http://www.math.uah.edu/stat/dist/Joint.html . See also the figure at the beginning of
https://en.wikipedia.org/wiki/Marginal_distribution .

By multiplying entries from the two marginals with each other and comparing to the joint entry, you can decide about independence of the marginals.

Intuitively: **Does the joint distribution contain information that cannot be inferred from the marginals alone? If the marginals are independent, the answer is "no".**

Conditional distribution

Look again at the figure at the beginning of the Wikipedia page "marginal distribution".

The marginal is the sum of the joint probabilities in a row (column).

The conditional distribution given $Y = y_1$ ($P(X|Y = y_1)$) is the first row interpreted as its own distribution, i.e. normalized to sum=1. (Therefore the formula $P(X, Y) = P(X|Y) * P(Y)$, which corrects the normalization again by multiplying with the marginal probability).

What does a discrete bivariate distribution look like?

What does a continuous bivariate distribution look like?

What does a discrete 3-dimensional distribution look like?

What does a continuous 3-dimensional distribution look like?

Univariate and bivariate Gaussian

Density estimation

Data and histograms, and density estimation:
From a data-analytic viewpoint, the difference between continuous and discrete variables is smaller than one would think. We may think of data being generated, e.g. by a Gaussian distribution, which is a continuous distribution. However, the data is just individual measurements. Automatically, we make a histogram. This models the data as generated by a probability mass function, but this is pure fiction. What bin size should we choose for the histogram? Don't let the fact that R will select parameters for you deceive you! Drawing a histogram is a first modeling decision.

What we are going to cover:

- Networks for modeling biological relationships
- Introduce/refresh some basic concepts of probability
- **Regression and correlation**
- Information theory
- Partial correlation and the inverse variance-covariance matrix
- Markov random-fields and Gaussian Graphical Models
- Bayes Networks

Butte AJ, . . . , Kohane IS (2000) Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. PNAS

"Comprehensive pair-wise correlations were calculated between gene expression and measures of agent susceptibility. Associations weaker than a threshold strength were removed, leaving networks of highly correlated genes and agents called relevance networks."

The correlation coefficient

$$Cov(X, Y) = E(X, Y) - E(X)E(Y)$$

$$\rho = Cov(X, Y)/\sigma_X \sigma_Y$$

See
`http://www.math.uah.edu/stat/sample/Covariance.html`

Correlation and independence

Why do we use a correlation coefficient of (near) 0 as an indicator of independence?
Let data be drawn from a bivariate Gaussian. Then, if their correlation coefficient is 0, this means that our estimate of the covariance in the bivariate Gaussian Σ is 0.

Correlation coefficient and linearity

Commonly, we think of the correlation coefficient as measuring linearity in a relationship between two variables. However, here we made the assumption that the two variables stem from a bivariate Gaussian. What is the connection?

Difference between the population correlation coefficient and the sample correlation. Sample correlation coefficient is an estimator for population correlation coefficient. Remember: It is easy to estimate the correlation coefficient from data (if only we have enough data).

On the other hand: When the two variables do not come from a bivariate Gaussian, they may have correlation coefficient 0 and yet this does not tell us anything. See examples in Wikipedia `https://en.wikipedia.org/wiki/Pearson_correlation_coefficient` . Still, the theorem gets frequently applied to samples where we are not so sure about their distributional origin.

What we are going to cover:

- Networks for modeling biological relationships
- Introduce/refresh some basic concepts of probability
- Regression and correlation
- **Information theory**
- Partial correlation and the inverse variance-covariance matrix
- Markov random-fields and Gaussian Graphical Models
- Bayes Networks

http://s3.amazonaws.com/academia.edu.documents/
30706964/butte.pdf?AWSAccessKeyId=
AKIAIWOWYYGZ2Y53UL3A&Expires=1497778675&Signature=
wrtfMPjnsvtAce2gKSchxqx3jk4%3D&
response-content-disposition=inline%3B%20filename%
3DMutual_information_relevance_networks_fu.pdf

Information theory

Given a discrete RV X with density $p(x_i)$.

Entropy H:

$$H(X) = \sum p(x_i) \log_2 p(x_i)$$

Given 2 discrete RVs X, Y with densities $p(x_i)$ and $p(y_i)$.
Def: Relative entropy, aka Kulback-Leibler divergence

$$H(X|Y) = \sum p(x_i) \log_2 \frac{p(x_i)}{p(y_i)}$$

This is 0 when $p(x) = p(y)$, but it is asymmetric!

Now, let $p(x_i, y_j)$ be the joint distribution.
Def: Mutual information

$$I(X;Y) = \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

Interpretation of this formula

Mutual information is the relative entropy between the joint distribution and the product-of-marginals.

Theorem

$I(X;Y) = 0$ if and only if X and Y are independent.

See also the wikipedia page "Mutual Information". There one also finds the formulas for continuous distributions (with integrals instead of sums).

Example: Look again at the Figure at the beginning of the Wikipedia page marginal distribution. We want to know whether the two marginal distributions are independent. So we compute their MI. First compute the product of the marginals, which is a new 2D histogram. To obtain the MI, we take the difference between the given 2D distribution and the product-of-marginals. This also explains why the theorem holds: We said before that the distributions are independent when the product of the marginals is equal to the 2D histogram. In this case the relative entropy would be 0, and vice versa.

Estimating mutual information?

To compute information, entropy, rel. entropy, MI we need a (continuous or discrete) distribution. However, while for the correlation coefficient there is a population correlation coefficient and a sample correlation, for entropy there is no canonical definition of a "sample entropy". Mostly estimation of entropy relies on binning (=drawing a histogram) of sample data. See next week's seminar presentation.

Correlation coefficient and mutual information

Let X, Y be drawn from a bivariate Gaussian:

$$I(X;Y) = \frac{1}{2} * log(1 - \rho^2)$$

(https://en.wikipedia.org/wiki/Mutual_information#Linear_correlation)

Contingency tables, chi-squared test

If data is given in a contingency table.
χ^2 test compares expected and observed number of observations.
Advantage: knowledge of the distribution.
See wikipedia entries "chi-squared test" and "contingency table".

Conditional independence

See Bishop, Section 8.2

Given 3 RVs A, B, C: A and B are conditionally independent given C iff:

$$P(A|B, C) = P(A|C)$$

i.e. A does not depend on B, but only on C. Or:

$$P(A, B|C) \stackrel{\text{chain rule}}{=} P(A|B, C)P(B|C) \stackrel{\text{cond.ind.}}{=} P(A|C)P(B|C)$$

This is the chain rule from above because $P(A, B|C) := \frac{P(A,B,C)}{P(C)}$.

One can visualize these things in Rubick's cube (without colors, but frequencies instead).