Computational Systems Biology: Biology X

Bud Mishra

Room 1002, 715 Broadway, Courant Institute, NYU, New York, USA

L#8:(November-08-2010)
Cancer and Signals

Outline

Outline

Multicellularity

- In a multicellular organism, a group of cells must work together to accomplish a particular "function."
- No single cell can perform the entire function, but only its "component" of the function: **action**.
- The appropriate **action** depends upon the global state: microenvironment, stress, oxygen, pH, etc.
- No single cell may know the global state: but only some "component" of the state: **type**.

Sender-Receiver Game

- A sender cell or ECM (extra-cellular matrix) knows the type, and based on it sends a subset of few available **signals**.
- A receiver cell receives the **signals** and activates kinases, transcriptional factors to turn on certain genes to perform certain **actions**.
- Sender wants the signals to carry as much information as possible, and specific actions to be carried out as a result of the signals.
- Receiver wishes the signals to encode the global sate as best as possible, and the actions to confirm to the state as informatively as possible.

Signaling

- Intracrine (within a cell)
- Autocrine (originating from the same cell)
- Paracrine (originating from nearby cells)
- Endocrine (system-wide)

Signal

- Growth Factors (Kinases)
- Motility (Integrin)
- Apoptosis (Caspases)
- Metabolism (Hypoxia, Anoxia, etc.)
- Autophagy
- Metaplasia (Transdifferentiation, Dedifferentiation)
- Meta-signals (Mutators?)

Outline

Information theory

- Information theory is based on probability theory (and statistics).
- **Basic concepts**: *Entropy* (the information in a random variable) and *Mutual Information* (the amount of information in common between two random variables).
- The most common unit of information is the **bit** (based log 2). Other units include the **nat**, and the **hartley**.

Entropy

- The entropy H of a discrete random variable X is a measure of the amount uncertainty associated with the value X.
- Suppose one transmits 1000 bits (0s and 1s). If these bits are known ahead of transmission (to be a certain value with absolute probability), logic dictates that no information has been transmitted. If, however, each is equally and independently likely to be 0 or 1, 1000 bits (in the information theoretic sense) have been transmitted.

Entropy

- Between these two extremes, information can be quantified as follows.
- If **X** is the set of all messages x that X could be, and $p(x)$ is the probability of X given x, then the **entropy of** X is defined as

$$H(x) = E_X[I(x)] = -\sum_{x \in X} p(x) \log p(x).$$

Here, $I(x)$ is the self-information, which is the entropy contribution of an individual message, and E_X is the expected value.

- An important property of entropy is that it is maximized when all the messages in the message space are equiprobable $p(x) = 1/n$, i.e., most unpredictable, in which case $H(X) = \log n$.
- The binary entropy function (for a random variable with two outcomes $\in \{0, 1\}$ or $\in \{H, T\}$:

$$H_b(p, q) = -p \log p - q \log q, \quad p + q = 1.$$

Joint entropy

- The joint entropy of two discrete random variables X and Y is merely the entropy of their pairing: $\langle X, Y \rangle$.
- Thus, if X and Y are independent, then their joint entropy is the sum of their individual entropies.

$$H(X, Y) = E_{X,Y}[-\log p(x, y)] = -\sum_{x,y} \log p(x, y).$$

- For example, if (X,Y) represents the position of a chess piece Ñ X the row and Y the column, then the joint entropy of the row of the piece and the column of the piece will be the entropy of the position of the piece.

Conditional Entropy or Equivocation

- The conditional entropy or conditional uncertainty of X given random variable Y (also called the equivocation of X about Y) is the average conditional entropy over Y:

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

- A basic property of this form of conditional entropy is that:

$$H(X|Y) = H(X,Y) - H(Y).$$

Mutual Information (Transinformation)

- Mutual information measures the amount of information that can be obtained about one random variable by observing another.
- The mutual information of X relative to Y is given by:

$$I(X;Y) = E_{X,Y}[SI(x,y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$

where SI (**Specific mutual Information**) is the pointwise mutual information.

- A basic property of the mutual information is that

$$I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y) = I(Y;X).$$

  That is, knowing Y, we can save an average of $I(X;Y)$ bits in encoding X compared to not knowing Y. Note that mutual information is **symmetric**.
- It is important in communication where it can be used to maximize the amount of information shared between sent and received signals.

Kullback-Leibler Divergence (Information Gain)

- The Kullback-Leibler divergence (or information divergence, information gain, or relative entropy) is a way of comparing two distributions: a “true” probability distribution $p(X)$, and an arbitrary probability distribution $q(X)$.

$$
\begin{aligned}
D_{KL}(p(X)\|q(X)) &= \sum_{x\in X} p(x)\log\frac{p(x)}{q(x)} \\
&= \sum_{x\in X} [-p(x)\log q(x)] - [-p(x)\log p(x)]
\end{aligned}
$$

- If we compress data in a manner that assumes $q(X)$ is the distribution underlying some data, when, in reality, $p(X)$ is the correct distribution, the Kullback-Leibler divergence is the number of average additional bits per datum necessary for compression.
- Although it is sometimes used as a 'distance metric,' it is not a true metric since it is not symmetric and does not satisfy the triangle inequality (making it a semi-quasimetric).

- Mutual information can be expressed as the average Kullback-Leibler divergence (information gain) of the posterior probability distribution of X given the value of Y to the prior distribution on X:

$$\begin{aligned} I(X;Y) &= E_{p(Y)}[D_{KL}(p(X|Y=y)\|p(X)] \\ &= D_{KL}(p(X,Y)\|p(X)p(Y)). \end{aligned}$$

  In other words, mutual information $I(X,Y)$ is a measure of how much, on the average, the probability distribution on X will change if we are given the value of Y. This is often recalculated as the divergence from the product of the marginal distributions to the actual joint distribution.

- Mutual information is closely related to the log-likelihood ratio test in the context of contingency tables and the multinomial distribution and to Pearson's χ^2 test.

Source theory

- Any process that generates successive messages can be considered a source of information.
- A memoryless source is one in which each message is an independent identically-distributed random variable, whereas the properties of ergodicity and stationarity impose more general constraints. All such sources are stochastic.

Information Rate

- **Rate** Information rate is the average entropy per symbol. For memoryless sources, this is merely the entropy of each symbol, while, in the case of a stationary stochastic process, it is

$$r = \lim_{n\to\infty} H(X_n | X_{n-1}, X_{n-2} \dots)$$

- In general (e.g., nonstationary), it is defined as

$$r = \lim_{n\to\infty} \frac{1}{n} H(X_n, X_{n-1}, X_{n-2} \dots)$$

- In information theory, one may thus speak of the “rate” or “entropy” of a language.

Rate Distortion Theory

- $R(D) =$ Minimum achievable rate under a given constraint on the expected distortion.
- $X =$ random variable; $T =$ alphabet for a compressed representation.
- If $x \in X$ is represented by $t \in T$, there is a distortion $d(x, t)$

$$\begin{aligned} R(D) &= \min_{\{p(t|x):\langle d(x,t)\rangle \leq D\}} I(T, X). \\ \langle d(x,t)\rangle &= \sum_{x,t} p(x,t) d(x,t) \\ &= \sum_{x,t} p(x) p(t|x) d(x,t) \end{aligned}$$

- Introduce a Lagrange multiplier parameter β and
- Solve the following **variational problem**

$$\mathcal{L}_{min}[p(t|x)] = I(T;X) + \beta \langle d(x,t) \rangle_{p(x)p(t|x)}.$$

- We need

$$\frac{\partial \mathcal{L}}{\partial p(t|x)} = 0.$$

Since

$$\mathcal{L} = \sum_x p(x) \sum_t p(t|x) \log \frac{p(t|x)}{p(t)} + \beta \sum_x p(x) \sum_t p(t|x) d(x,t),$$

we have

$$p(x) \left[\log \frac{p(t|x)}{p(t)} + \beta d(x,t) \right] = 0.$$

$$\Rightarrow \frac{p(t|x)}{p(t)} \propto e^{-\beta d(x,t)}.$$

Summary

- In summary,

$$p(t|x) = \frac{p(t)}{Z(x,\beta)} e^{-\beta d(x,t)} \qquad p(t) = \sum_x p(x) p(t|x).$$

$Z(x,\beta) = \sum_t p(t) \exp[-\beta d(x,t)]$ is a Partition Function.

- The Lagrange parameter in this case is positive; It is determined by the upper bound on distortion:

$$\frac{\partial R}{\partial D} = -\beta.$$

Redescription

- Some hidden object may be observed via two views X and Y (two random variables.)
- Create a common descriptor T
- Example $X =$ words, $Y =$ topics.

$$
\begin{aligned}
R(D) &= \min_{p(t|x):I(T:Y)\geq D} I(T;X) \\
\mathcal{L} &= I(T:X) - \beta I(T;Y)
\end{aligned}
$$

- Proceeding as before, we have

$$
\begin{aligned}
p(t|x) &= \frac{p(t)}{Z(x,\beta)} e^{-\beta D_{KL}[p(y|x) \| p(y|t)]} \\
p(t) &= \sum_x p(x) p(t|x) \\
p(y|t) &= \frac{1}{p(t)} \sum_x p(x,y) p(t|x) \\
p(y|x) &= \frac{p(x,y)}{p(x)}
\end{aligned}
$$

- **Information Bottleneck** $= T$.

Blahut-Arimoto Algorithm

- Start with the basic formulation for RDT; Can be changed *mutatis mutandis* for IB.
- **Input:** $p(x)$, T, and β
- **Output:** $p(t|x)$

Step 1. Randomly initialize $p(t)$

Step 2. **loop until** $p(t|x)$ converges (to a fixed point)

Step 3. $p(t|x) := \frac{p(t)}{Z(x,\beta)} e^{-\beta d(x,t)}$

Step 4. $p(t) := \sum_x p(x) p(t|x)$

Step 5. **endloop**

Convex Programming: Optimization of a convex function over a convex set $\mapsto$ Global optimum exists!

[End of Lecture #8]