Computational functional genomics
(Spring 2001: Lecture 5)

David K. Gifford
(Adapted from a lecture by Tommi S. Jaakkola)
MIT LCS and AI Lab

gifford@mit.edu
Topics

- Significance testing
  - statistical tests, test statistics
  - p-values, power of a test

- Maximum a posterior estimates
  - Bayes’ rule
  - Dirichlet distributions

- Statistical models
  - linear regression

- Normalization with spiked controls
Statistical tests: example

- Defining the hypothesis:
  Let $X_1$ and $X_2$ are the random variables corresponding to the expression levels of the two genes.
  
  The null hypothesis $H_0$: $X_1$ and $X_2$ are uncorrelated:
  
  $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right)$ \hspace{1cm} (1)

  The alternative hypothesis $H_1$: $X_1$ and $X_2$ can be correlated:
Statistical tests: example

- Defining the hypothesis:

Let $X_1$ and $X_2$ are the random variables corresponding to the expression levels of the two genes

The null hypothesis $H_0$: $X_1$ and $X_2$ are uncorrelated:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right)$$ \hspace{1cm} (2)

The alternative hypothesis $H_1$: $X_1$ and $X_2$ can be correlated:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$ \hspace{1cm} (3)

where $\Sigma_{ij}$ is the covariance between $X_i$ and $X_j$ ($\sigma_i^2 = \Sigma_{ii}$)
Statistical tests: example

- The alternative hypothesis \( H_1 \) is more expressive in terms of explaining the observed data.

- We need to find a way of testing whether this difference is significant.
Test statistic

- Likelihood ratio statistic

\[ T(X^{(1)}, \ldots, X^{(n)}) = 2 \log \frac{P(X^{(1)}, \ldots, X^{(n)} | \hat{H}_1)}{P(X^{(1)}, \ldots, X^{(n)} | \hat{H}_0)} \quad (4) \]

Larger values of \( T \) imply that the model corresponding to the null hypothesis \( H_0 \) is much less able to account for the observed data.

- To evaluate the P-value, we also need to know the sampling distribution for the test statistic.

In other words, we need to know how the test statistic \( T(X^{(1)}, \ldots, X^{(n)}) \) varies if the null hypothesis \( H_0 \) is correct.
Test statistic cont’d

- For the likelihood ratio statistic, the sampling distribution is $\chi^2$ with degrees of freedom equal to the difference in the number of free parameters in the two hypotheses.

- Once we know the sampling distribution, we can compute the P-value

\[ p = \text{Prob}(T(X^{(1)}, \ldots, X^{(n)}) \geq T_{\text{obs}} | H_0) \]  

(5)
Degrees of freedom

- How many degrees of freedom do we have in the two models?

\[ H_0 : \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) \]

\[ H_1 : \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \]
Degrees of freedom

• How many degrees of freedom do we have in the two models?

\[ H_0 : \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \right) \]

\[ H_1 : \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \]

• The observed data overwhelmingly supports \( H_1 \)
Maximum likelihood principle: Binomial

- Maximum likelihood principle: we find the parameter \( \hat{\theta} \) that maximize the likelihood of the observed data

\[
\hat{\theta} = \arg\max_{\theta} L(x^{(1)}, \ldots, x^{(n)}|\theta)
\]  

(6)

The Maximum likelihood estimate (MLE) for the Binomial PMF is

\[
P(k_N|\theta) = \binom{N}{k} \theta^k (1 - \theta)^{N-k}
\]  

(7)

\[
\log P(k_N|\theta) = \log \binom{N}{k} + k \log \theta + (N - k) \log (1 - \theta)
\]  

(8)

\[
\frac{d}{d\theta} P(k_N|\theta) = \frac{k}{\theta} - \frac{N - k}{1 - \theta}
\]  

(9)

\[
0 = \frac{k}{\theta} - \frac{N - k}{1 - \theta}
\]  

(10)

\[
\hat{\theta} = \frac{k}{N}
\]  

(11)
Bayes’ Rule

- Key to Bayesian analysis is Bayes’ Rule

\[ P(A, B) = P(A|B)P(B) = P(B|A)P(A) \quad (12) \]
\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (13) \]
Maximum a Posterior Estimators (MAP)

- Assume that we know something about a coin before we observe $N$ trials
- Prior knowledge can take on many forms
  - Assumptions (mRNA levels are never negative)
  - Data (other experiments suggests that protein A regulates gene B)
  - Estimates (our best estimate of the parameters so far)
- How do we express this knowledge so that it can be used in a principled way?
- Represent this knowledge as a distribution over model parameters
  - In the case of a coin, as a distribution over $\theta$
Maximum a Posterior Estimators (MAP)

- Bayesians use prior knowledge when analyzing data
  - This can lead to different conclusions from the same data, depending on your prior
- Frequentists believe that conclusions from data should always be the same
- Using Bayes’ Rule in our Binomial example:

\[
P(\theta|k_N) = \frac{P(k_N|\theta)P(\theta)}{P(k_N)} \tag{14}
\]

- Let’s represent \( P(\theta) \) as:

\[
P(\theta) = C(\alpha)\theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1} \tag{15}
\]

\[
\alpha_1 = pS + 1 \tag{16}
\]

\[
\alpha_2 = (1-p)S + 1 \tag{17}
\]
Dirichlet Distributions

• $P(\theta)$ is a Dirichlet distribution, and is a conjugate distribution to the Binomial distribution:

\[
P(\theta) = C(\alpha)\theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1}
\]

\[
\alpha_1 = pS + 1
\]

\[
\alpha_2 = (1-p)S + 1
\]

• This binomial form of the Dirichlet distribution is called the Beta distribution.

• Now:

\[
P(\theta|k_N) = \frac{\binom{N}{k} C(\alpha)\theta^k pS (1-\theta)^{(N-k)+(1-p)S}}{P(k_N)}
\]

\[
\frac{d}{d\theta} P(\theta|k_N) = \frac{k + pS}{\theta} - \frac{(N-k) + (1-p)S}{1-\theta}
\]

\[
\hat{\theta}_{MAP} = \frac{k + pS}{N + S}
\]