Lecture 33. Phylogeny methods, part 5 (Likelihood methods)

Joe Felsenstein

Department of Genome Sciences and Department of Biology

Likelihoods and odds ratios

Bayes' Theorem relates prior and posterior probabilities of an hypothesis H:

$$\begin{aligned} \text{Prob}\,(H|D) &= \text{Prob}\,(H \text{ and } D) / \text{Prob}\,(D) \\ &= \text{Prob}\,(D|H)\,\text{Prob}\,(H) / \text{Prob}\,(D) \end{aligned}$$

The ratios of posterior probabilities of two hypotheses, H_1 and H_2 can be written, putting this into its "odds ratio" form ($\text{Prob}\,(D)$ cancels):

$$\frac{\text{Prob}\,(H_1|D)}{\text{Prob}\,(H_2|D)} = \frac{\text{Prob}\,(D|H_1)}{\text{Prob}\,(D|H_2)}\,\frac{\text{Prob}\,(H_1)}{\text{Prob}\,(H_2)}$$

Note that this says that the posterior odds in favor of H_1 over H_2 are the product of prior odds and a likelihood ratio. The likelihood of the hypothesis H is the probability of the observed data given it, $\text{Prob}\,(D|H)$. This is *not* the same as the probability of the hypothesis given the data. That is the posterior probability of H and requires that we also have a believable prior probability $\text{Prob}\,(H)$

Rationale of likelihood inference

If the data consists of n items that are conditionally independent given the hypothesis H_i,

$$\begin{aligned} & \text{Prob}\ (D|H_i) \\ = \ & \text{Prob}\ (D^{(1)}|H_i)\ \text{Prob}\ (D^{(2)}|H_i) \ldots\ \text{Prob}\ (D^{(n)}|H_i). \end{aligned}$$

and we can then write the likelihood ratio $\text{Prob}\ (D|H_1)\ /\ \text{Prob}\ (D|H_2)$ as a product of ratios:

$$\frac{\text{Prob}\ (D|H_1)}{\text{Prob}\ (D|H_2)} = \left(\prod_{i=1}^{n} \frac{\text{Prob}\ (D^{(i)}|H_1)}{\text{Prob}\ (D^{(i)}|H_2)} \right)$$

If the amount of data is large the likelihood ratio terms will dominate and push the result towards the correct hypothesis. This can console us somewhat for the lack of a believable prior.

Properties of likelihood inference

Likeihood inference has (usually) properties of

- Consistency. As the number of data items n gets large, we converge to the correct hypothesis with probability 1.
- Efficiency. Asymptotically, the likelihood estimate has the smallest possible variance (it need not be best for any finite number n of data points).

A simple example – coin tossing

If we toss a coin which has heads probability p and get HHTTHTHHTTT the likelihood is

$$\begin{aligned} L &= \text{Prob}\ (D|p) \\ &= pp(1-p)(1-p)p(1-p)pp(1-p)(1-p)(1-p) \\ &= p^5(1-p)^6 \end{aligned}$$

so that trying to maximize it we get

$$\frac{dL}{dp} = 5p^4(1-p)^6 - 6p^5(1-p)^5$$

finding the ML estimate

and searching for a value of p for which the slope is zero:

$$\frac{dL}{dp} = p^4(1-p)^5\left(5(1-p) - 6p\right) = 0$$

which has roots at $p = 0$, $p = 1$, and $p = 5/11$

Log likelihoods

Alternatively, we could maximize not L but its logarithm. This turns products into sums:

$$\ln L \; = \; 5 \ln p + 6 \ln(1 - p)$$

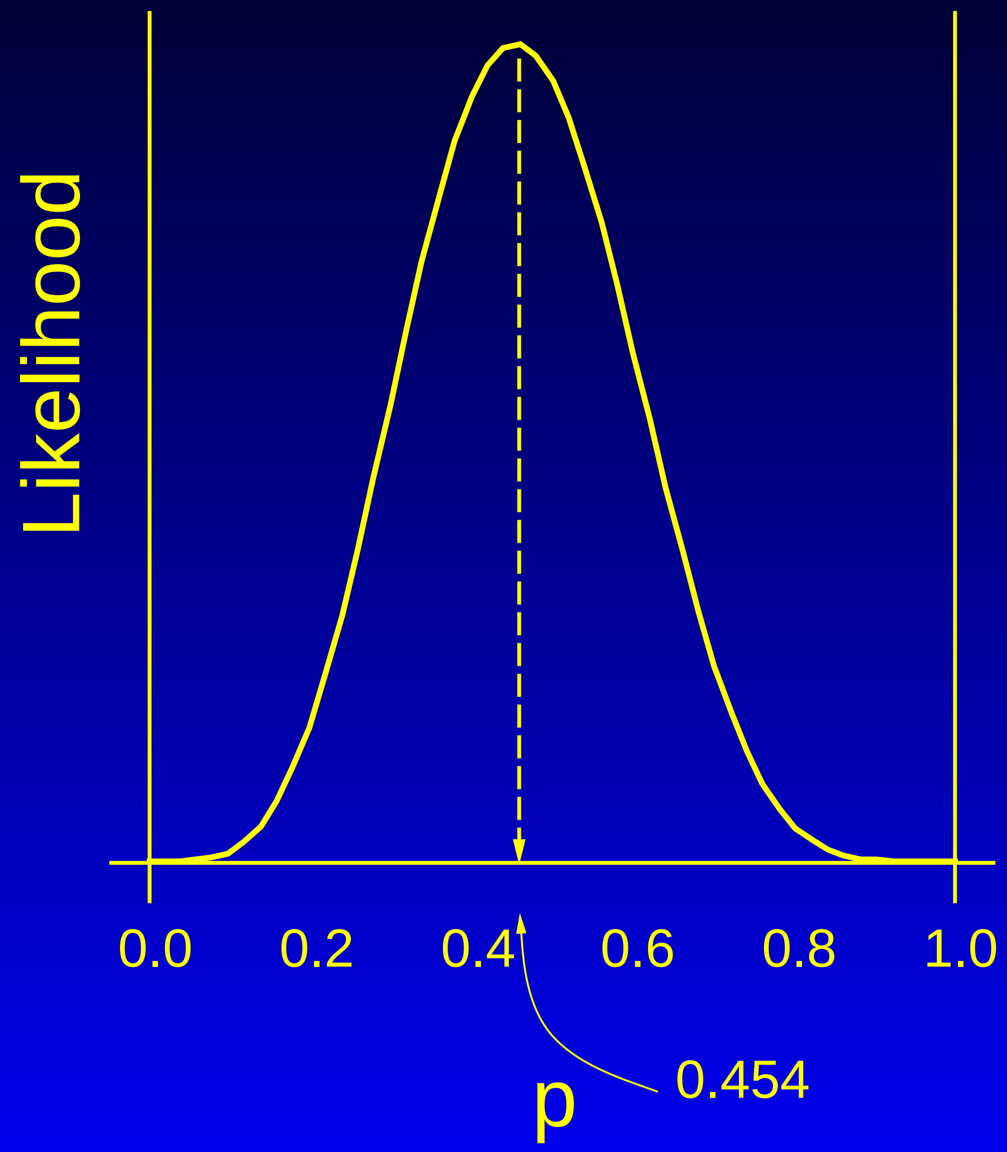
whereby

$$\frac{d(\ln L)}{dp} \; = \; \frac{5}{p} - \frac{6}{(1-p)} \; = \; 0$$

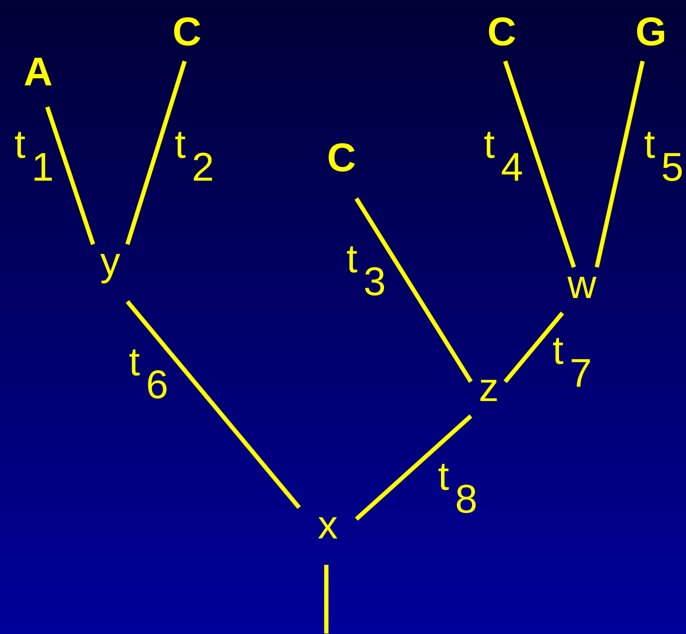
so that finally

$$\widehat{p} \; = \; 5/11$$

Likelihood curve for coin tosses



Likelihood on trees



A tree, with branch lengths, and the data at a single site

This example is used to describe calculation of the likelihood

Since the sites evolve independently on the same tree,

$$L = \text{Prob}\ (D|T) = \prod_{i=1}^{m} \text{Prob}\ \left(D^{(i)}|T\right)$$

Likelihood at one site on a tree

We can compute this by summing over all assignments of states x, y, z

and w to the interior nodes

$$\text{Prob}\ \left(D^{(i)}|T\right) =$$

$$\sum_x \sum_y \sum_z \sum_w \text{Prob}\ (A, C, C, C, G, x, y, z, w|T)$$

Computing the terms

For each combination of states, the Markov process allows us to express it as a product of probabilities of a series of changes, with the probability that we start in state x:

$$\text{Prob}\,(A, C, C, C, G, x, y, z, w|T) =$$

$$\text{Prob}\,(x) \quad \text{Prob}\,(y|x, t_6) \quad \text{Prob}\,(A|y, t_1)\ \text{Prob}\,(C|y, t_2)$$

$$\text{Prob}\,(z|x, t_8) \quad \text{Prob}\,(C|z, t_3)$$

$$\text{Prob}\,(w|z, t_7)\ \text{Prob}\,(C|w, t_4)\ \text{Prob}\,(G|w, t_5)$$

Computing the terms

Summing this up, there are $4^4 = 256$ terms in this case:

$$\text{Prob}\ \left(D^{(i)}|T\right) =$$

$$\sum_x \sum_y \sum_z \sum_w$$

$$\text{Prob}\ (x) \quad \text{Prob}\ (y|x,t_6) \quad \text{Prob}\ (A|y,t_1)\ \text{Prob}\ (C|y,t_2)$$

$$\text{Prob}\ (z|x,t_8) \quad \text{Prob}\ (C|z,t_3)$$

$$\text{Prob}\ (w|z,t_7)\ \text{Prob}\ (C|w,t_4)\ \text{Prob}\ (G|w,t_5)$$

Getting a recursive algorithm

This seems hopeless, but when we move the summation signs as far right as possible

$$\text{Prob}\ \left(D^{(i)}|T\right) =$$

$$\sum_x \text{Prob}\ (x)$$
$$\left(\sum_y \text{Prob}\ (y|x,t_6) \quad \text{Prob}\ (A|y,t_1)\ \text{Prob}\ (C|y,t_2) \right)$$
$$\left(\sum_z \text{Prob}\ (z|x,t_8) \quad \text{Prob}\ (C|z,t_3) \right.$$
$$\left. \left(\sum_w \text{Prob}\ (w|z,t_7)\ \text{Prob}\ (C|w,t_4)\ \text{Prob}\ (G|w,t_5) \right) \right)$$

The pruning algorithm

Note that the pattern of parentheses in the previous expression is the

$$(A, C)\,(C, (C, G))$$

If $L_k^{(i)}(s)$ is the probability of everything that is observed from node k on the tree on up, at site i, conditional on node k having state s, we can express

$$\left(\sum_w \text{Prob}\,(w|z, t_7)\ \text{Prob}\,(C|w, t_4)\ \text{Prob}\,(G|w, t_5) \right)$$

as:

$$\left(\sum_w \text{Prob}\,(w|z, t_7) L_7^{(i)}(w) \right)$$

and the algorithm is:

Continuing with this we find that the following algorithm computes the L_k's from the L_ℓ and L_m above them,

$$L_k^{(i)}(s) = \left(\sum_x \text{Prob}\ (x|s, t_\ell)\, L_\ell^{(i)}(x) \right) \times \left(\sum_y \text{Prob}\ (y|s, t_m)\, L_m^{(i)}(y) \right)$$

Starting and finishing the recursion

At the top of the tree the definition of the L's specifies that they look like this

$$\left(L^{(i)}(A), L^{(i)}(C), L^{(i)}(G), L^{(i)}(T)\right) = (1, 0, 0, 0)$$

and at the bottom the likelihood for the whole site can be computed simply by weighting by the equilibrium state probabilities

$$L^{(i)} = \sum_x \pi_x L_0^{(i)}(x)$$

Ambiguity and error in the sequences

Ambiguity. If a tip has an ambiguity state such as R (purine, either A or G) we use

$$L^{(i)} = (1, 0, 1, 0)$$

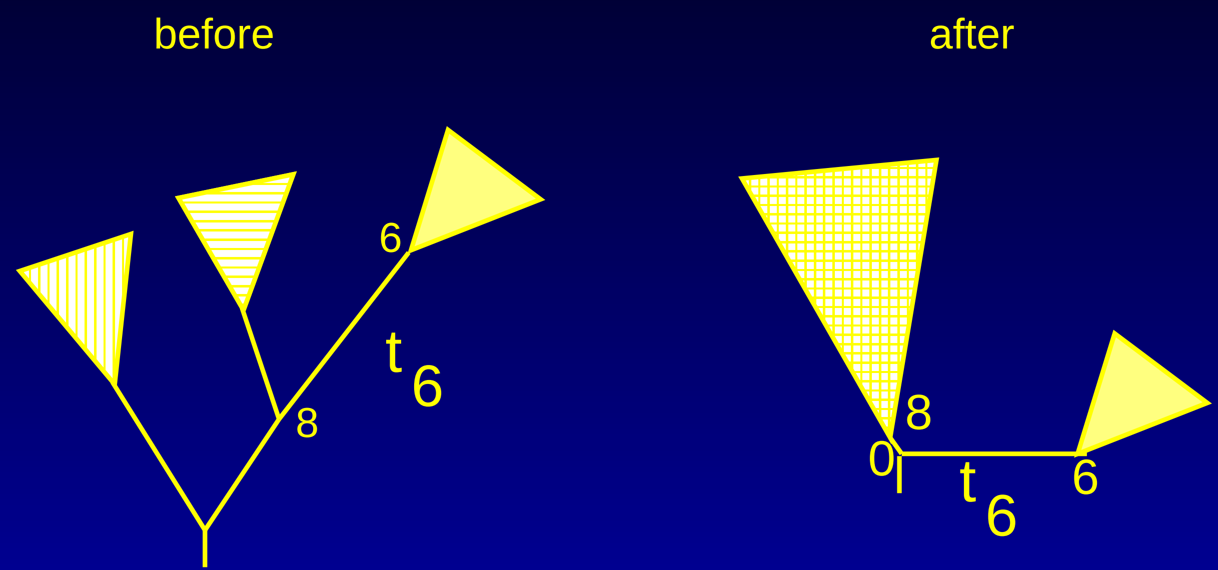
and if it has an unknown nucleotide (“N")

$$L^{(i)} = (1, 1, 1, 1)$$

This handles ambiguities naturally.

Error. If our sequencing has probability $1 - \varepsilon$ of finding the correct nucleotide, and $\varepsilon/3$ of inferring each of the three other possibilities, when an A is observed, the four values should be $(1 - \varepsilon,\ \varepsilon/3,\ \varepsilon/3,\ \varepsilon/3)$, and when a C is observed, they should be $(\varepsilon/3,\ 1 - \varepsilon,\ \varepsilon/3,\ \varepsilon/3)$
The result is a simple handling of sequencing error, provided it occurs independently in different bases.

The tree is effectively unrooted



The region around nodes 6 and 8 in the tree, when a new root (node 0) is placed in that branch
The subtrees are shown as shaded triangles

For the tree on the left of the figure above,

$$L^{(i)} = \sum_y \sum_z \sum_x \text{Prob}\,(x)\ \text{Prob}\,(y|x, t_6)\ \text{Prob}\,(z|x, t_8).$$

using reversibility ...

Reversibility of the substitution process guarantees us that

$$\text{Prob}\ (x)\ \text{Prob}\ (y|x, t_6) = \ \text{Prob}\ (y)\ \text{Prob}\ (x|y, t_6).$$

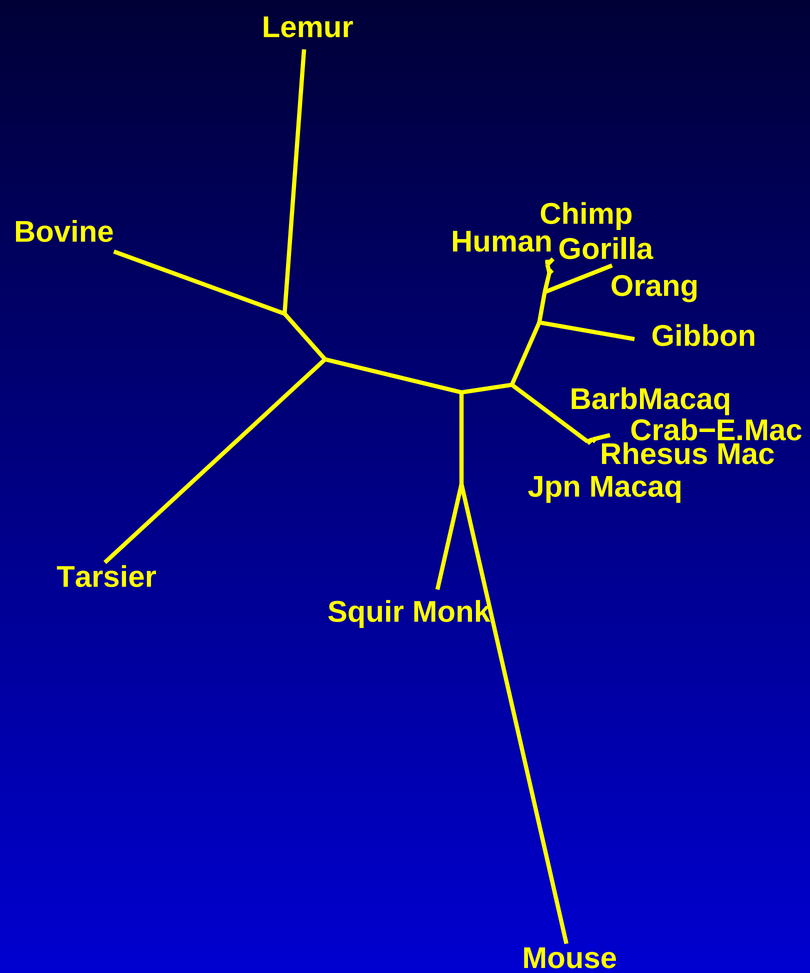
Substituting, we get

$$L^{(i)} = \sum_y \sum_z \sum_x \ \text{Prob}\ (y)\ \text{Prob}\ (x|y, t_6)\ \text{Prob}\ (z|x, t_8)$$

Finally we see that this is the same as the likelihood for a tree rooted at node 8:

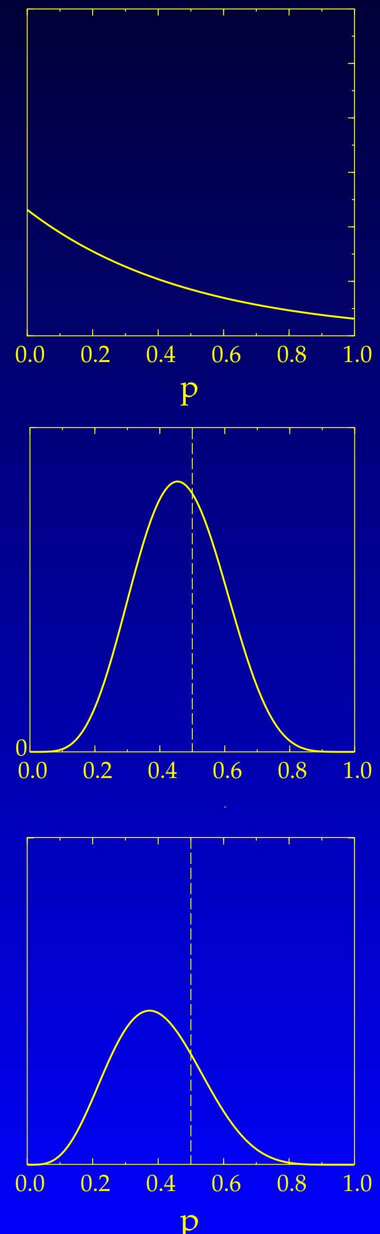
$$L_0^{(i)}(z) = L_8^{(i)}(z)\ \text{Prob}\ (z)\ \text{Prob}\ (w|z, t_6) L_6^{(i)}(w)$$

A numerical example



A 232-nucleotide mitochondrial noncoding region data set over 14 species gives this ML tree with $\ln L = -2616.86$ with a transition/transversion ratio of 30

Bayesian inference with coin tossing:



Bayesian methods

An example of Bayesian inference with coin-tossing. The probability of heads is assumed to have a prior (top) which is a truncated exponential with mean 0.34348 on the interval (0,1). The likelihood curve (middle) and the posterior on the probability of heads (bottom) are shown, when there are 11 tosses with 5 heads.

Bayesian phylogeny methods

Bayesian inference has been applied to inferring phylogenies (Rannala and Yang, 1996; Mau and Larget, 1997; Li, Pearl and Doss, 2000).

- All use a prior distribution on trees. The prior has enough influence on the result that its reasonableness should be a major concern. In particular, the depth of the tree may be seriously affected by the distribution of depths in the prior.
- All use Markov Chain Monte Carlo (MCMC) methods (we will introduce these in our discussion of coalescents) They sample from the posterior distribution.
- When these methods make sense they not only get you a point estimate of the phylogeny, they get you a distribution of possible phylogenies.

References

Barry, D., and J. A. Hartigan. 1987. Statistical analysis of hominoid molecular evolution. *Statistical Science* **2:** 191-210. **[ML with full 12-parameter model, estimated on each branch]**

Edwards, A. W. F., and L. L. Cavalli-Sforza. 1964. Reconstruction of evolutionary trees. pp. 67-76 in *Phenetic and Phylogenetic Classification*, ed. V. H. Heywood and J. McNeill. Systematics Association Publ. No. 6, London. **[first paper on likelihood for phylogenies]**

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17:** 368-376. **[Made likelihood practical for n species]**

Felsenstein, J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology* **22:** 240-249. **[The "pruning" algorithm]**

Fisher, R. A. 1912. On an absolute criterion for fitting frequency curves. *Messenger of Mathematics* **41:** 155-160. **[First modern paper introducing likelihood]**

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A* **222:** 309-368. **[Likelihood in generality]**

References

Kashyap, R. L., and S. Subas. 1974. Statistical estimation of parameters in a phylogenetic tree using a dynamic model of the substitutional process. *Journal of Theoretical Biology* **47:** 75-101. **[Second paper applying likelihood to molecular sequences]**

Li, S., D. Pearl, and H. Doss. 2000. Phylogenetic tree construction using Markov chain Monte Carlo. *Journal of the American Statistical Association* **95:** 493-508. **[Bayesian inference of phylogenies by MCMC]**

Mau, B., M. A. Newton, and B. Larget. 1997. Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Molecular Biology and Evolution* **14:** 717-724. **[Bayesian inference of phylogenies by MCMC]**

Neyman, J. 1971. Molecular studies of evolution: a source of novel statistical problems. pp. 1-27 in *Statistical Decision Theory and Related Topics*, ed. S. S. Gupta and J. Yackel. Academic Press, New York. **[First application of likelihood to molecular sequences]**

Rannala, B. and Z. Yang. 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J. Molecular Evolution* **43:** 304-311. **[Bayesian inference of phylogenies by MCMC]**

How it was done

This projection produced as a PDF, not a PowerPoint file, and viewed using the Full Screen mode (in the View menu of Adobe Acrobat Reader):

- using the `prosper` style in LaTeX,
- using Latex to make a `.dvi` file,
- using `dvips` to turn this into a Postscript file,
- using `ps2pdf` to mill it into a PDF file, and
- displaying the slides in Adobe Acrobat Reader.

Result: nice slides using freeware.