Week 4: Consistency; History / philosophy, distance methods

Genome 570

January, 2014
The counterexample

The long branches have probabilities of change $p$, the short branches have probabilities of change $q$. This is the canonical case of “long branch attraction”.

Week 4: Consistency; History / philosophy, distance methods – p.2/63
Pattern probabilities

If tip pattern is 1100, and internal nodes are 1, 1 the probability is

$$\frac{1}{2}(1 - p)(1 - q)(1 - q)pq$$

and in general summing over all four possibilities:

$$P_{1100} = \frac{1}{2}((1 - p)(1 - q)^2pq + (1 - p)^2(1 - q)^2q + p^2q^3 + pq(1 - p)(1 - q)^2)$$
Pattern probabilities

\[ P_{xxyy} = (1 - p)(1 - q)[q(1 - q)(1 - p) + q(1 - q)p] \]
\[ + pq[(1 - q)^2(1 - p) + q^2p] \]

\[ P_{xyxy} = (1 - p)q[q(1 - q)p + q(1 - q)(1 - p)] \]
\[ + p(1 - q)[p(1 - q)^2 + (1 - p)q^2] \]

\[ P_{xxyx} = (1 - p)q[(1 - p)q^2 + p(1 - q)^2] \]
\[ + p(1 - q)[q(1 - q)p + q(1 - q)(1 - p)] \]
Taking differences

\[ P_{xyxy} - P_{xxyx} = (1 - 2q) \left[ q^2(1 - p)^2 + (1 - q)^2 p^2 \right] \]

Which is always positive as long as \( q < 1/2 \) and either \( p \) or \( q \) is positive. Thus \( P_{xyxy} > P_{xxyx} \) so we don’t need to concern ourselves with \( P_{xxyx} \).

To have \( P_{xxyy} \) be the largest of the three, we only need to know that

\[ P_{xxyy} - P_{xyxy} > 0 \]

and after a struggle that turns out to require

\[ (1 - 2q) \left[ q(1 - q) - p^2 \right] > 0 \]

which (provided \( q < 1/2 \)) is true if and only if

\[ q(1 - q) > p^2 \]
Pattern frequencies win out

![Graph showing pattern frequencies](image)

- Pattern 1: \(x\) \(x\) \(y\) \(y\), frequency 0.112
- Pattern 2: \(x\) \(y\) \(x\) \(y\), frequency 0.098
- Pattern 3: \(y\) \(x\) \(y\) \(x\), frequency 0.076

Number of characters vs. Frequencies of the three patterns

Week 4: Consistency; History / philosophy, distance methods – p.6/63
Conditions for inconsistency

\[ p^2 < q(1-q) \]

consistent

inconsistent
Example for patterns with DNA

1(C)  p  3(A)
 \downarrow \quad q \quad \downarrow
 q \quad p \quad q
Z   w
2(C)  \quad q \quad 4(A)

root
Calculating a pattern frequency

\[
\text{Prob} \ [\text{CCAA}] = \frac{1}{18} (1 - p)(1 - q)^2 pq + \frac{1}{27} pq^2 (1 - p)(1 - q) + \frac{1}{162} p^2 q^2 (1 - q) + \frac{7}{972} p^2 q^3 + \frac{1}{12} (1 - p)^2 (1 - q)^2 q.
\]
Pattern frequencies

\[
\begin{align*}
\text{Prob } [\text{xxyy}] &= (1 - p)^2 q(1 - q)^2 + \frac{2}{3} p(1 - p)q(1 - q)^2 \\
&+ \frac{4}{9} p(1 - p)q^2(1 - q) + \frac{2}{27} p^2q^2(1 - q) \\
&+ \frac{7}{81} p^2q^3 \\
\text{Prob } [\text{xyxy}] &= \frac{1}{3} (1 - p)^2q^2(1 - q) + \frac{2}{9} p(1 - p)q^2(1 - q) \\
&+ \frac{4}{27} p(1 - p)q^3 + \frac{1}{3} p^2(1 - q)^3 \\
&+ \frac{2}{9} p^2q^2(1 - q) + \frac{2}{81} p^2q^3 \\
\text{Prob } [\text{xyyx}] &= \frac{1}{81} (1 - p)^2q^3 + \frac{2}{3} p(1 - p)q(1 - q)^2 \\
&+ \frac{4}{9} p(1 - p)q^3 + \frac{1}{9} p^2q(1 - q)^2 \\
&+ \frac{6}{27} p^2q^2(1 - q) + \frac{2}{81} p^2q^3.
\end{align*}
\]
Conditions for inconsistency with DNA

\[ p < \frac{-18q + 24q^2 + \sqrt{243q - 567q^2 + 648q^3 - 288q^4}}{9 - 24q + 32q^2} \]

(This will *not* be on the test).

For small \( p \) and \( q \) this is approximately

\[ \frac{1}{3}p^2 < q \]
Conditions for inconsistency with DNA

- Inconsistent
- Consistent
Inconsistency with a clock

Cases like this were discovered by Michael Hendy and David Penny in 1989.
Example showing inconsistency with a clock

- **E**
  - **A**
    - **B**
      - **C**
        - **D**

Trees correct out of 1000

- **b = 0.03**
- **b = 0.02**
History and philosophy

Issues:
- How did work on numerical methods for phylogenies get started?
- What is the logical basis of inferring phylogenies?
- How does all that relate to classification?
Example showing inconsistency with a clock

- E
  - A
    - B
      - C
      - D

Trees correct out of 1000

- b = 0.03
- b = 0.02

Sites

Week 4: Consistency; History / philosophy, distance methods – p.16/63
Ernst Mayr and George Gaylord Simpson


- Major figures in the completion of the “modern synthesis” or “Neodarwinian synthesis” in the 1940s.

- Leaders of the “evolutionary systematics” approach to taxonomic classification, dominant until the 1970s.
Evolutionary-systematic classification

A pattern of grades with very unequal rates of overall evolution is implicit in the use of paraphyletic groups in Mayr and Simpson’s practice.
A horse tree drawn by Simpson

Note the “fat” branches which are somewhat ambiguous. Are trees starting to dissolve?
A phylogeny of the living Craniata

... as an example of groups that are in the traditional classification system but may or may not be monophyletic.
Vertebrates are a monophyletic group
Reptiles and fishes are paraphyletic groups

... and for that matter amphibians are iffy too, if this tree is right.
Positions on classification as of about 1960

- **Evolutionary systematics.** George Gaylord Simpson and Ernst Mayr led a movement that allowed non-monophyletic (paraphyletic) groups such as reptiles, on the assumption that groups could be separated by real differences of rates of evolution (sometimes “grades” rather than “clades”).

- **Phylogenetic systematics.** Willi Hennig advocated purely monophyletic classification.

- **Phenetics.** Sokal and Sneath advocated making a classification without reference to evolution, using numerical clustering methods.
Technological change post World War II

- Former physicists found molecular biology (first protein sequence, 1951)
- Former codebreakers and atomic bomb builders build the early computers (first stored-program digital computer, 1949)
- Most U.S. universities got their first computer about 1957.
- First sequences of same gene in multiple species in late 1950s.
Molecular evolution gets off the ground

Zuckerkandl and Pauling in 1962 discussed using trees to infer ancestral sequences, and named this “chemical paleogenetics”. They were about 30 years ahead of their time.

Emile Zuckerkandl in 1986

Linus in 1962, from Nobel Peace Prize web page

(But then it isn’t fair to anyone to compare them unfavorably to Linus Pauling).
Peter Sneath in 1962 and Robert Sokal in 1964

... as they were advocating phenetic classification. Sokal did pioneering investigations of parsimony methods – intending to show that they wouldn’t work well!
A tree of bees (Michener is the world’s greatest bee systematist). Michener was the one who wanted to interpret this as a phylogeny. It was inferred by a clustering method (not, as misstated in my book, by UPGMA).
Cavalli-Sforza and Edwards, 1963; Edwards, 1970

The picture on the left was taken by the famous population geneticist Motoo Kimura when he and his family visited Pavia and were taken on an excursion to a vineyard.
The first phylogeny by parsimony

Gene frequencies of blood group polymorphisms. This is by minimum length in a space of gene frequencies.
That tree drawn out on a map

Note that the lineages from the Northwest Coast going down to Polynesia are dependent on where on the map the splits are placed and that is somewhat arbitrary.
Camin in the 1970s, one of the Caminalcules

Camin noticed (in 1965) that students who did the best job recovering the true “phylogeny” of the Caminalcules made the reconstruction which required the fewest changes of state.
J. S. Farris and Arnold Kluge in the 1980s

Margaret Dayhoff, 1966

- A major pioneer of molecular databases (starting in 1965)
- (With Richard Eck) made the first numerical phylogenies using molecular data
- Presented trees organized by gene families in the *Atlas of Protein Sequences* (later the PIR database) in 1966.
- Compiled the first empirical substitution rate matrices for amino acids, intended to form the basis of a probabilistic model of protein evolution.
Walter Fitch (1929-2011):

- The first major distance matrix method (1967)
- Developed algorithm (1971) that counts changes of state in DNA parsimony.
- Introduced the terms and concepts of orthology and paralogy.
- Co-founded the journal MBE and the society SMBE.
Fitch and Margoliash’s 1967 distance tree

This is for cytochrome C sequences. It is the first distance matrix method published, if you don’t count clustering methods.
But who are the two guys standing next to Charlie Cantor? (Hint: one has a Nobel Prize, the other is a member of my own department). Cantor later made important technical discoveries in genomics. Jukes was a nutritional biochemist who was the primary person responsible for insisting that pregnant women get folic acid in their diet.
A major figure in mathematical statistics, he and Egon Pearson invented the Neyman-Pearson testing machinery. Statisticians don’t even recall that he once worked on maximum likelihood inference of phylogenies from protein sequences (usually he was known as a pointed critic of likelihood).
Willi Hennig (1913-1976) and Walter Zimmermann (1890-1980)

Zimmermann pioneered the method later advocated and made important by Hennig. Hennig was the major advocate of a purely monophyletic classification system.
Hennig and conflict of characters

W. Hennig (1966) says that in the case of homoplasy,

“it becomes necessary to recheck the interpretation of [the] characters"

He also says (1966, p. 121)

“the more certainly characters interpreted as apomorphous (not characters in general) are present in a number of different species, the better founded is the assumption that these species form a morphological group."

Farris, Kluge, and Eckardt (1970) argue that this should be translated as:

“the more characters certainly interpretable as apomorphous ..."
Did Hennig invent parsimony?

Farris (1983, p. 8):

I shall use the term in the sense I have already mentioned: most parsimonious genealogical hypotheses are those that minimize requirements for ad hoc hypotheses of homoplasy. If minimizing ad hoc hypotheses is not the only connotation of "parsimony" in general usage, it is scarcely novel. Both Hennig (1966) and Wiley (1975) have advanced ideas closely related to my usage. Hennig defends phylogenetic analysis on the grounds of his auxiliary principle, which states that homology should be presumed in the absence of evidence to the contrary. This amounts to the precept that homoplasy ought not be postulated beyond necessity, that is to say parsimony.
Hennig’s auxiliary principle

Hennig discusses the case in which “only one character can certainly or with reasonable probability be interpreted as apomorphic.”

Hennig (1966, p. 121):

In such cases it is impossible to decide whether the common character is indeed synapomorphous or is to be interpreted as parallelism, homoiology, or even as convergence. I have therefore called it an “auxiliary principle” that the presence of apomorphic characters in different species “is always reason for suspecting kinship [i.e. that the species belong to a monophyletic group], and that their origin by convergence should not be assumed a priori” (Hennig 1953). This was based on the conviction that “phylogenetic systematics would lose all the ground on which it stands" if the presence of apomorphic characters in different species were considered first of all as convergences (or parallelisms), with proof to the contrary required in each case.

This is usually considered to be a statement of parsimony. Is it?
Farris and Kluge on Hennig and parsimony

Unfortunately, AIV is not sufficiently detailed to allow us to select a unique criterion for choosing a most preferable tree. We know that trees on which the monophyletic groups share many steps are preferable to trees on which this is not so. But AIV deals only with single monophyletic groups and does not tell us how to evaluate a tree consisting of several monophyletic groups. One widely used criterion – parsimony – could be used to select trees. This would be in accord with AIV, since on a most parsimonious tree OTUs [tips] that share many states (this is not the same as the OTUs’ being described by many of the same states) are generally placed together. We might argue that the parsimony criterion selects a tree most in accord with AIV by “averaging" in some sense the preferability of all the monophyletic groups of the tree. Other criteria, however, may also agree with AIV.

Farris, Eckardt and Kluge, 1970
Philosophical frameworks: hypothetico-deductive

Gaffney (1979, pp. 98-99)

In any case, in a hypothetico-deductive system, parsimony is not merely a methodological convention, it is a direct corollary of the falsification criterion for hypotheses (Popper, 1968a, pp. 144-145). When we accept the hypothetico-deductive system as a basis for phylogeny reconstruction, we try to test a series of phylogenetic hypotheses in the manner indicated above. If all three of the three possible three-taxon statements are falsified at least once, the least-rejected hypothesis remains as the preferred one, not because of an arbitrary methodological rule, but because it best meets our criterion of testability. In order to accept an hypothesis that has been successfully falsified one or more times, we must adopt an *ad hoc* hypothesis for each falsification .... Therefore, in a system that seeks to maximize vulnerability to criticism, the addition of *ad hoc* hypotheses must be kept to a minimum to meet this criterion.
more from Gaffney

Gaffney (1979)

“the use of derived character distributions as articulated by Hennig (1966) appears to fit the hypothetico-deductive model best.”

Gaffney (1979, p. 98)

“it seems to me that parsimony, or Ockham’s razor, is equivalent to ‘logic’ or ‘reason’ because any method that does not follow the above principle would be incompatible with any kind of predictive or causal system.”
Hypothetico-deductivists on falsification

Eldredge and Cracraft (1980, p. 69) are careful to point out that

“Falsified" implies that the hypotheses are proven false, but this is not the meaning we (or other phylogenetic systematists) wish to convey. It may be that the preferred hypothesis will itself be “rejected" by some synapomorphies.


In other words, we have no external criterion to say that a particular conflicting character is actually an invalid test. Therefore, saying that it is an invalid test simply because it is unparsimonious is a statement that is, itself, an ad hoc statement. With no external criterion, we are forced to use parsimony to minimize the total number of ad hoc hypotheses (Popper, 1968a: 145). The result is that the most parsimonious of the various alternates is the most highly corroborated and therefore preferred over the less parsimonious alternates.
Farris on hypothetico-deductivism

Farris (1983, p. 8):

Wiley [(1975)] discusses parsimony in a Popperian context, characterizing most parsimonious genealogies as those that are least falsified on available evidence. In his treatment, contradictory character distributions provide putative falsifiers of genealogies. As I shall discuss below, any such falsifier engenders a requirement for an ad hoc hypothesis of homoplasy to defend the genealogy. Wiley’s concept is then equivalent to mine.
Philosophical frameworks: Logical-parsimony

Beatty and Fink (1979):

We can account for the necessity of parsimony (or some such consideration) because evidence considerations alone are not sufficient. But we have no philosophical or logical argument with which to justify the use of parsimony considerations – a not surprising result, since this issue has remained a philosophical dilemma for hundreds of years.

This moves away from considering parsimony as justified by hypothetico-deductive frameworks and invokes it as its own justification.
Kluge and Wolf on logical parsimony

Kluge and Wolf (1993, p. 196):

Finally, we might imagine that some of the popularity of the aforementioned methodological strategies and resampling techniques, and assumption of independence in the context of taxonomic congruence and the cardinal rule of Brooks and McLennan (1991), derives from the belief that phylogenetic inference is hypothetico-deductive (e.g. Nelson and Platnick, 1984: 143-144), or at least that it should be. Even the uses to which some might put cladograms, such as “testing” adaptation (Coddington, 1988), are presented as hypothetico-deductive. But this ignores an alternative, that cladistics, and its uses, may be an abductive enterprise (Sober, 1988). We suggest that the limits of phylogenetic systematics will be clarified considerably when cladists understand how their knowledge claims are made (Rieppel, 1988; Panchen, 1992).

Again, moving away from hypothetico-deductivism.
Elliot Sober on falsification

Sober (1988, p. 126):

Popper’s philosophy of science is very little help here, because he has little to say about *weak* falsification. Popper, after all is a hypothetico-*deductivist*. For him, observational claims are deductive consequences of the hypothesis under test .... Deductivism excludes the possibility of probabilistic testing. A theory that assigns probabilities to various possible observational outcomes cannot be strongly falsified by the occurrence of any of them. This, I suggest, is the situation we confront in testing phylogenetic hypotheses. (AB)C is logically consistent with all possible character distributions (polarized or not), and the same is true of A(BC). [Emphasis in the original]

Sober sees Popper’s hypothetico-deductive framework as ill-suited to inference of phylogenies.
Philosophical foundations: Logical probability?

Popper’s corroboration formula

\[ C(h, e, b) = \frac{p(e, hb) - p(e, b)}{p(e, hb) - p(eh, b) + p(e, b)} \]

where
- \( b \) = background knowledge
- \( h \) = hypothesis
- \( e \) = evidence (= d, data?)

\[ C(h, e, b) = \frac{\text{Prob} (d|h) - \text{Prob} (d)}{\text{Prob} (d|h) - \text{Prob} (d&h) + \text{Prob} (d)} \]

This is used as a justification for parsimony, not actually as a statistical method in spite of people like Kluge calling it “logical probability”. Issue: how to compute \( \text{Prob} (d) \) if you aren’t a Bayesian and can’t weight by a prior on hypotheses?
Criticisms of statistical inference

Farris (1983, p.17):

The statistical approach to phylogenetic inference was wrong from the start, for it rests on the idea that to study phylogeny at all, one must first know in great detail how evolution has proceeded.

Kluge (1997a)

“As an aside, the fact that the study of phylogeny is concerned with the discovery of historical singularities means that calculus probability and standard (Neyman-Pearson) statistics cannot apply to that historical science ...."

If, after tossing a coin multiple times, you lose the coin, does he think you can’t then analyze those data?
Positions on classification nowadays

- **Phylogenetic systematics.** Willi Hennig advocated purely monophyletic classification. Now the (strongly) dominant approach.

- **Evolutionary systematics.** Has almost faded away. Its adherents were reluctant to make it algorithmic.

- **Phenetics.** Although Sokal and Sneath strongly influenced the field of numerical clustering, their approach to biological classification has few adherents.

- **IDMVM** One person (me) takes the view that It Doesn’t Matter Very Much, as we use the phylogeny, and, given that, we never use the classification system. This is widely regarded as a marginal crackpot view [“A bizarre thumb in the eye to systematists” – Michael Sanderson].

The predicted distances are the sums of branch lengths between those two species.
A tree and a set of two-species trees

The two-species trees correspond to the pairwise distances observed between the pairs of species. The tree also predicts two-species trees, because clipping off all other species you are left simply with the path between that pair of species.
A tree with branch lengths

Distance matrix methods always infer trees that have branch lengths, and they assume models of change of the characters (sequences).
Distance matrix methods

Each possible tree (with branch lengths) predicts pairwise distances

A B C D E F
A  0 10 9 12 16 9
B  10 0 10 6 9 9
C  9 10 0 10 15 2
D  12 6 10 0 6 13
E  16 9 15 6 0 15
F  9 9 2 13 15 0

observed distances calculated from the sequence data

Find the tree which comes closest to predicting the observed pairwise distances

Week 4: Consistency; History / philosophy, distance methods – p.56/63
The math of least squares trees

\[ Q = \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} (D_{ij} - d_{ij})^2 \]

\[ d_{ij} = \sum_{k} x_{ij,k} v_k \]

\[ Q = \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} \left( D_{ij} - \sum_{k} x_{ij,k} v_k \right)^2. \]

\[ \frac{dQ}{dv_k} = -2 \sum_{i=1}^{n} \sum_{j \neq i} w_{ij} x_{ij,k} \left( D_{ij} - \sum_{k} x_{ij,k} v_k \right) = 0. \]

Equations to solve to infer branch lengths by least squares on a given tree topology, which specifies the \( x_{ij,k} \).
Solving for least squares branch lengths

\[
\begin{align*}
D_{AB} + D_{AC} + D_{AD} + D_{AE} &= 4v_1 + v_2 + v_3 + v_4 + v_5 + 2v_6 + 2v_7 \\
D_{AB} + D_{BC} + D_{BD} + D_{BE} &= v_1 + 4v_2 + v_3 + v_4 + v_5 + 2v_6 + 3v_7 \\
D_{AC} + D_{BC} + D_{CD} + D_{CE} &= v_1 + v_2 + 4v_3 + v_4 + v_5 + 3v_6 + 2v_7 \\
D_{AD} + D_{BD} + D_{CD} + D_{DE} &= v_1 + v_2 + v_3 + 5v_4 + v_5 + 2v_6 + 3v_7 \\
D_{AE} + D_{BE} + D_{CE} + D_{DE} &= v_1 + v_2 + v_3 + v_4 + 4v_5 + 3v_6 + 2v_7 \\
D_{AC} + D_{AE} + D_{BC} + D_{BE} + D_{CD} + D_{DE} &= 2v_1 + 2v_2 + 3v_3 + 2v_4 + v_5 + 6v_6 + 4v_7 \\
D_{AB} + D_{AD} + D_{BC} + D_{BE} + D_{CD} + D_{DE} &= 2v_1 + 3v_2 + 3v_4 + 2v_5 + 4v_6 + 6v_7
\end{align*}
\]
A vector of all distances, stacked

\[ d = \begin{bmatrix}
D_{AB} \\
D_{AC} \\
D_{AD} \\
D_{AE} \\
D_{BC} \\
D_{BD} \\
D_{BE} \\
D_{CD} \\
D_{CE} \\
D_{DE}
\end{bmatrix} \]

They are in lexicographic (dictionary) order.
The design matrix and least squares equations

\[ X = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\end{bmatrix} \]

\[ X^T D = (X^T X) v. \]

So solution of equations is

\[ v = (X^T X)^{-1} X^T D \]
A diagonal matrix of weights

\[ W = \begin{bmatrix}
  w_{AB} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & w_{AC} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & w_{AD} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & w_{AE} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & w_{BC} & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & w_{BD} & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & w_{BE} & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{CD} & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{CE} & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & w_{DE} & 0
\end{bmatrix} \]
**Weighted least squares equations**

\[ X^T WD = (X^T WX) v, \]

\[ v = (X^T WX)^{-1} X^T WD. \]

These matrix equations solve for weighted least squares estimates of the branch lengths on a given tree. The tree topology is specified by the design matrix \( X \) and the weights are the elements of the diagonal matrix \( W \).
A statistical justification for least squares

$$SSQ = \sum_{i=1}^{n} \sum_{j \neq i} \frac{(D_{ij} - E(D_{ij}))^2}{\text{Var}(D_{ij})}.$$ 

This least squares method

- ... is what we would get by standard statistical least squares approaches \textit{if} the distances were normally distributed, independently varying, and had expectation and variance as shown

- ... but they actually aren’t independent in almost all cases (such as molecular sequences), but ...

- ... it can be shown that the estimate of the tree will be a consistent estimate in the case of non-independence, just not as efficient