

Independent Contrasts

Fig. 5 from Felsenstein (1985)

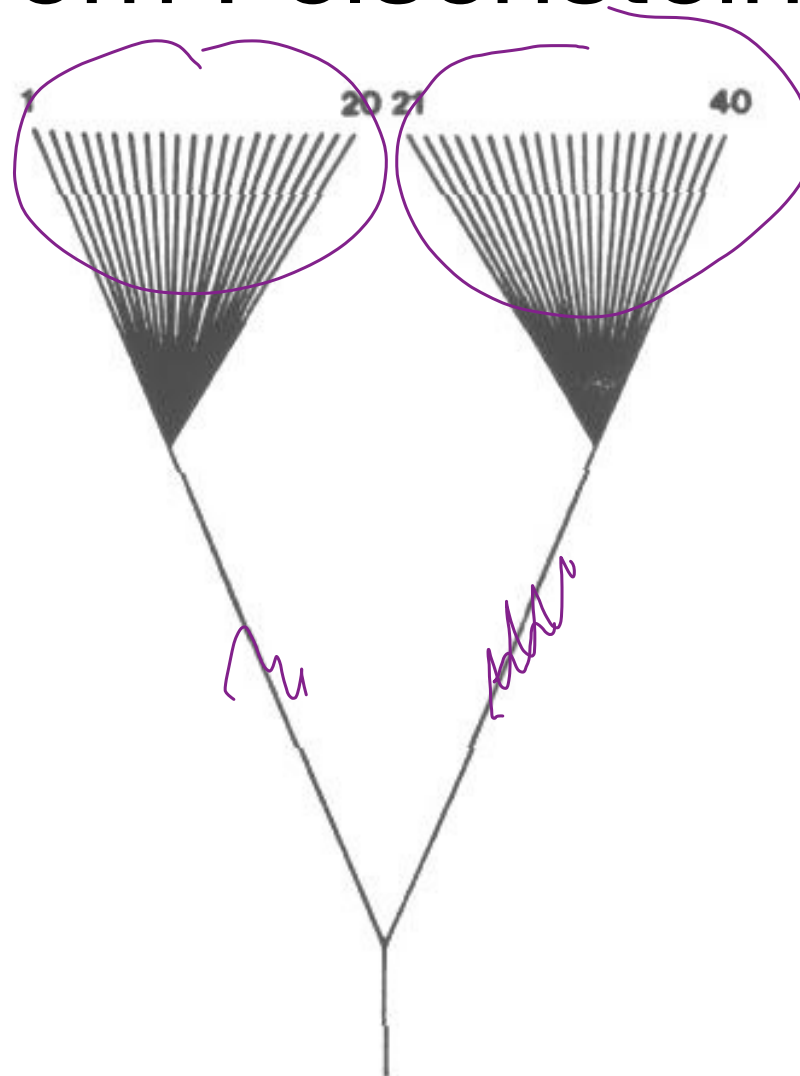


FIG. 5.—A “worst case” phylogeny for 40 species, in which there prove to be 2 groups each of 20 close relatives.

Felsenstein 1985b

Fig. 6 from Felsenstein (1985b)

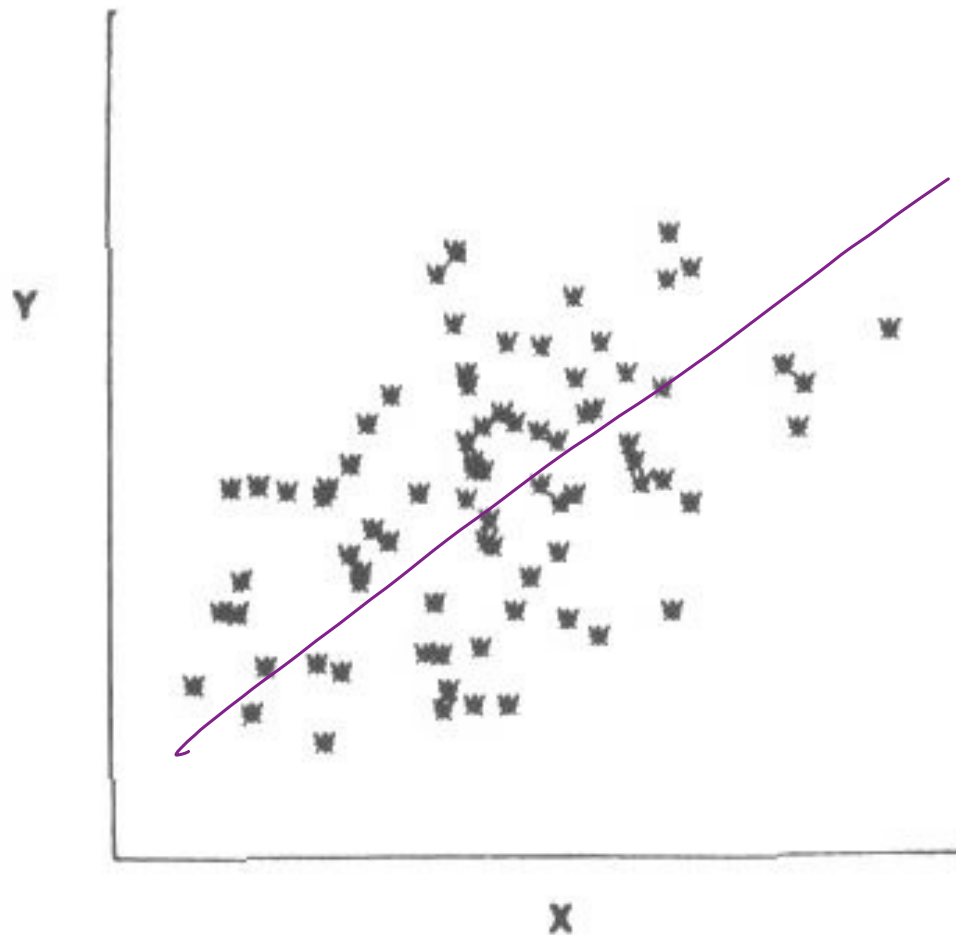


FIG. 6.—A typical data set that might be generated for the phylogeny in fig. 5 using the model of independent Brownian motion (normal increments) in each character.

Fig. 7 from Felsenstein (1985b)

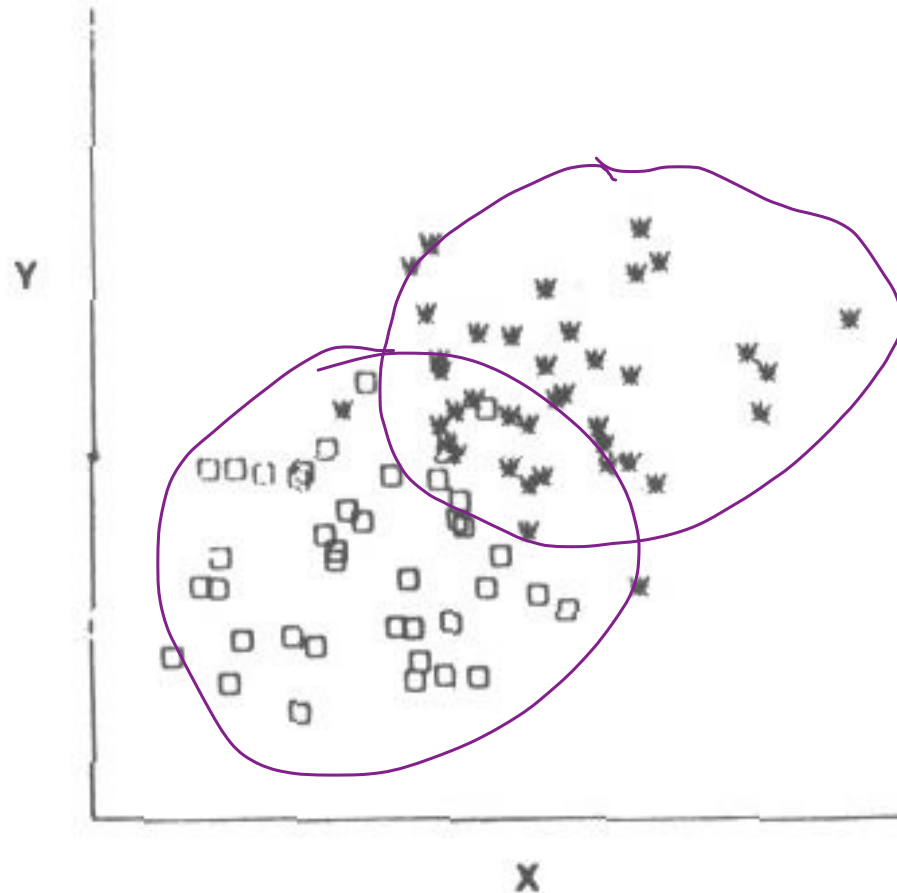
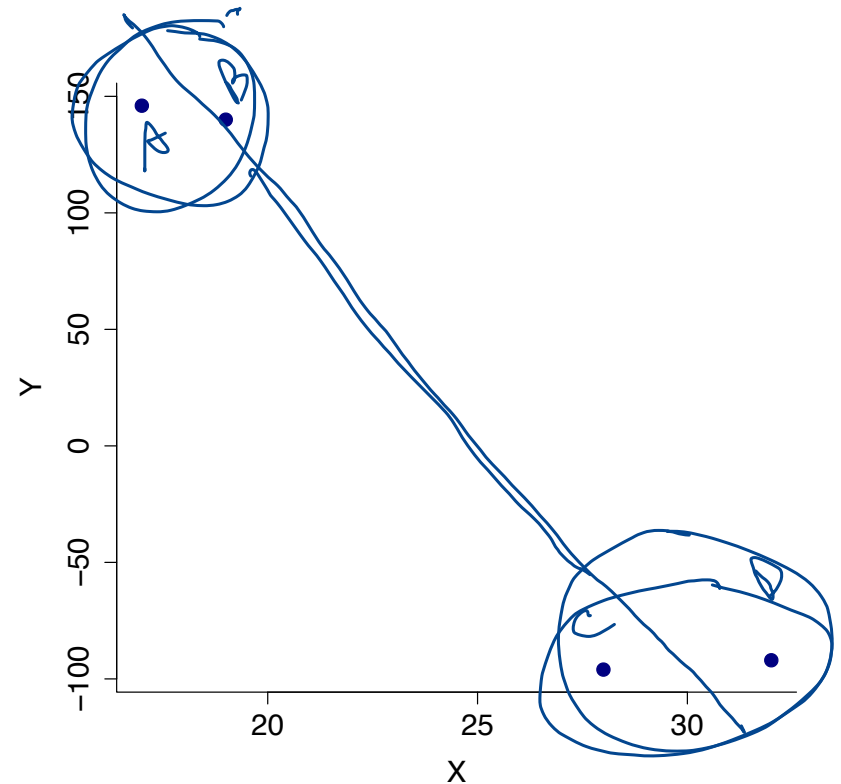


FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.

Data for two characters, X and Y

	X	Y
A	28	-96
B	32	-92
C	19	140
D	17	146

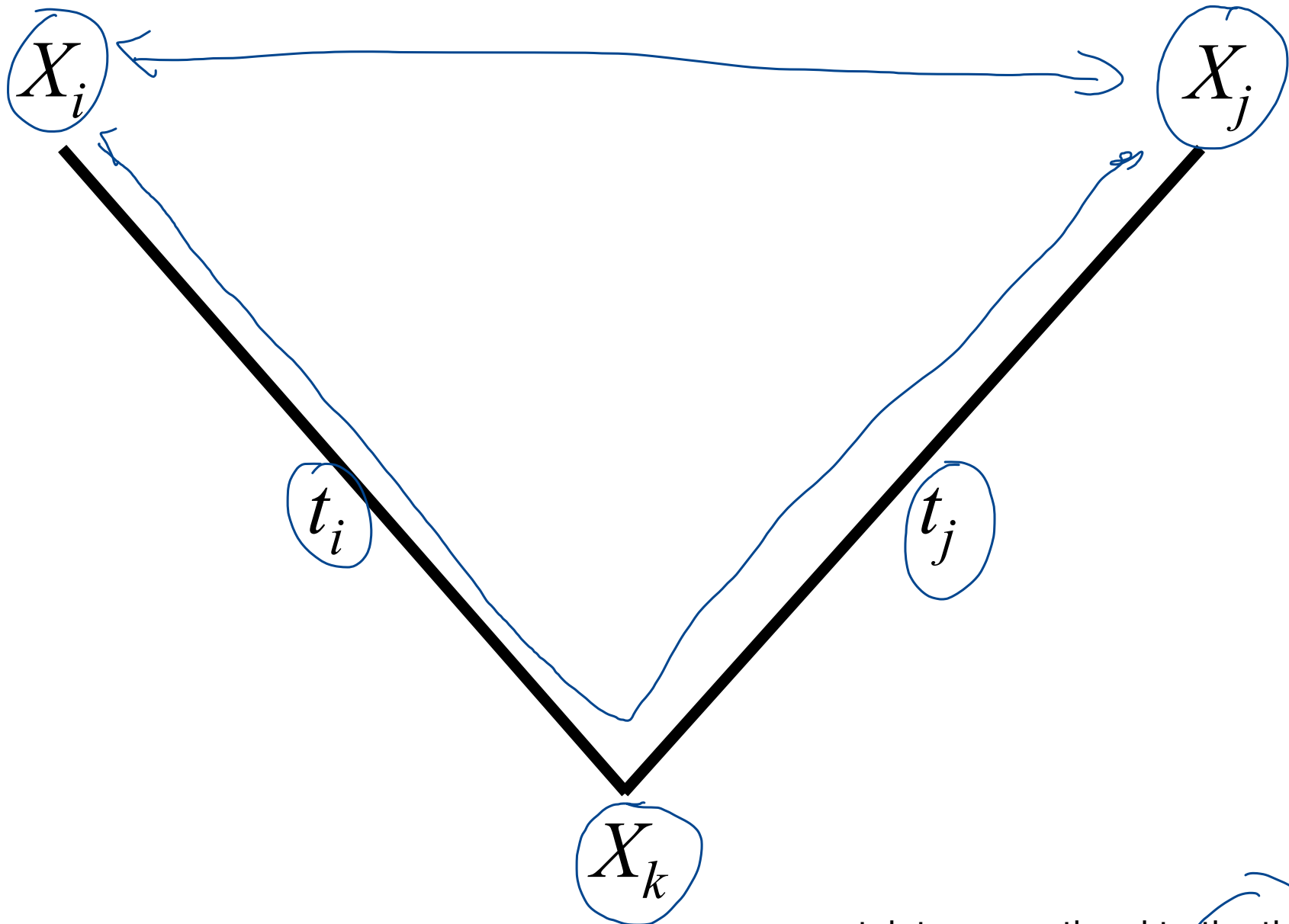


The negative correlation is evident in the plot and is significant, but would it weaken if it were recognized that there are not really 4 independent data points here...

Var(X) =	51	t = -5.29
Var(Y) =	18731	df = 2
Cov(XY) =	-867	P = 0.034
Correlation =	-0.947	

```
cor.test(c(28, 32, 19, 17), c(-96, -92, 140, 146), alternative="two.sided", method="pearson")
```

Brownian motion model





<https://plewis.github.io/applets/ou/>

T to toggle (select Brownian Motion)
S to simulate again
↑ or ↓ to change variance per unit time

Note that **uncertainty** in the distance separating two lineages increases with time

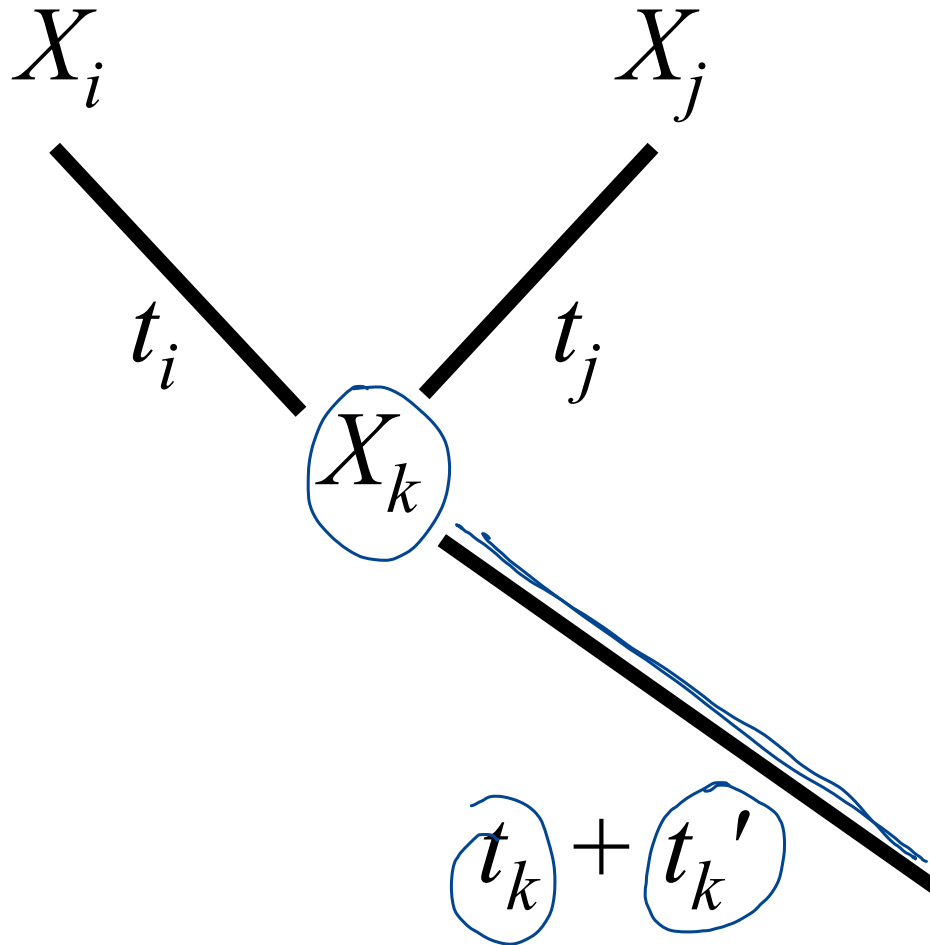
Brownian motion model

The diagram shows a V-shaped tree with two branches. The left branch is labeled t_i and ends at node X_i . The right branch is labeled t_j and ends at node X_j . The two branches meet at a common node labeled X_k . Below the tree, the equation for X_k is given as a weighted average of X_i and X_j . The equation is:

$$X_k = \frac{X_i \left(\frac{1}{t_i} \right) + X_j \left(\frac{1}{t_j} \right)}{\frac{1}{t_i} + \frac{1}{t_j}}$$

Blue circles highlight X_i , X_j , and the denominator $\frac{1}{t_i} + \frac{1}{t_j}$. Blue arrows point from the branch labels t_i and t_j to the corresponding terms in the equation.

Brownian motion model

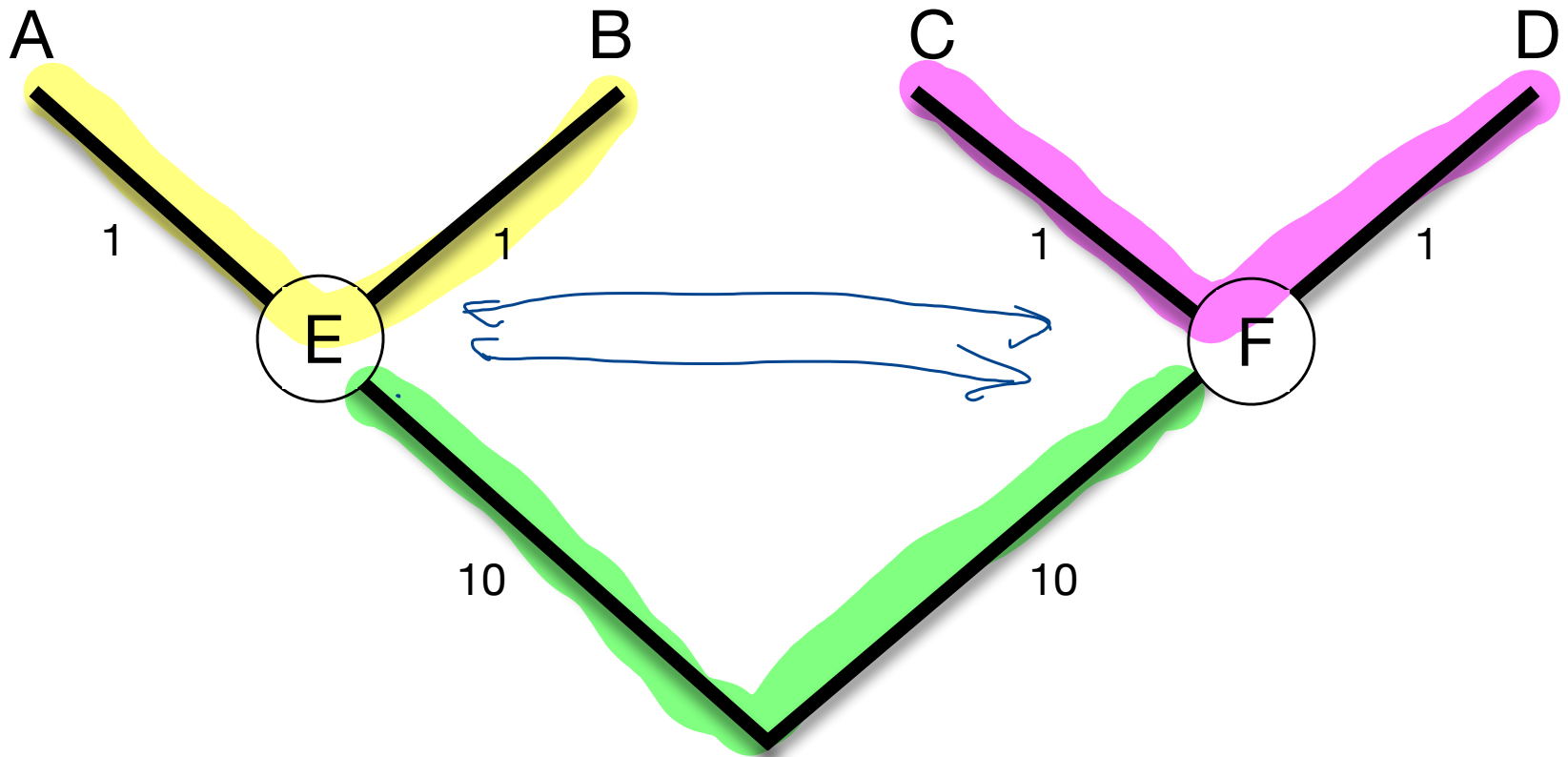


The amount of extra uncertainty that should be added to X_k is:

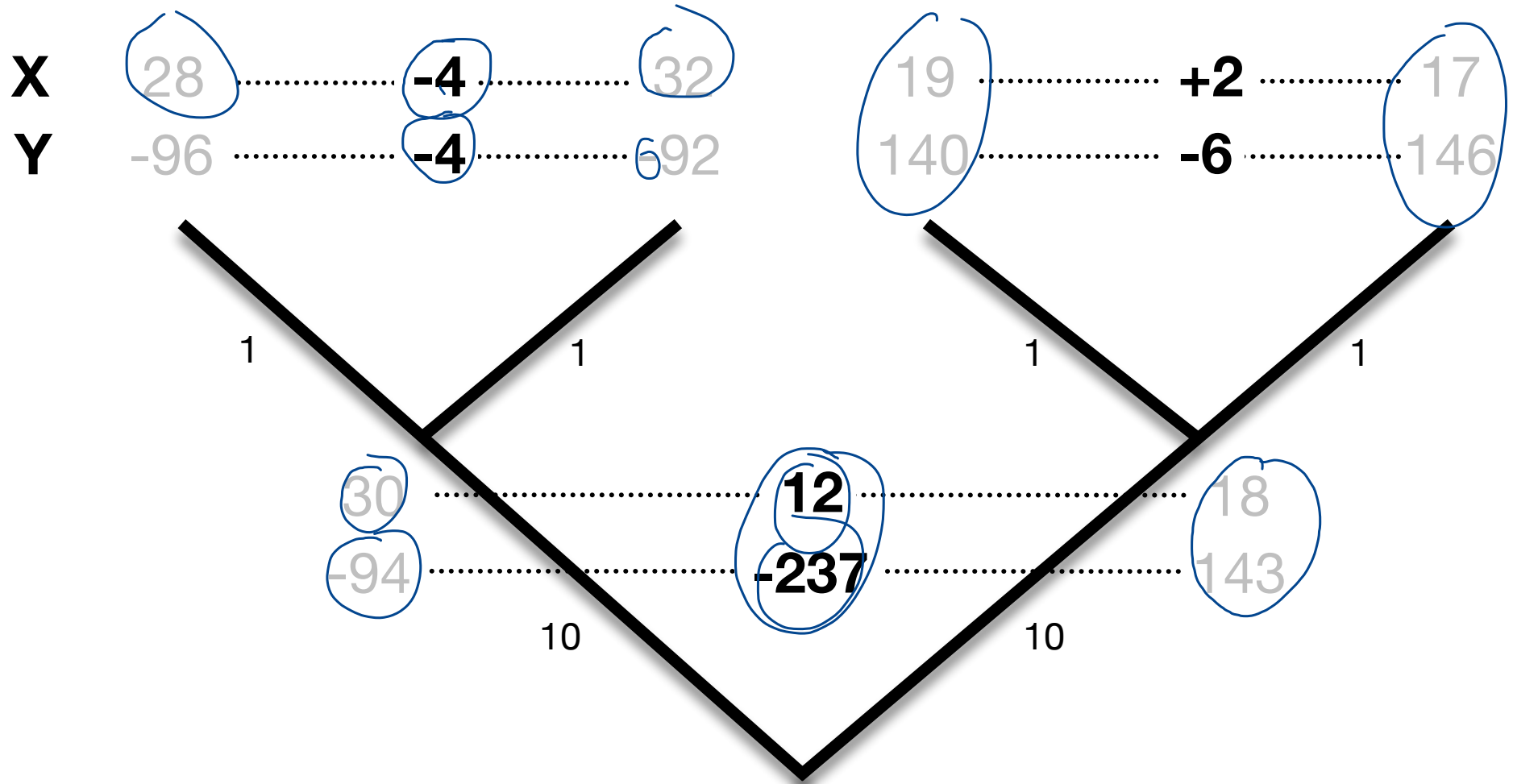
$$t'_k = \frac{t_i t_j}{t_i + t_j}$$

Data for two characters on tree

X 28 ←————→ 32 19 ←————→ 17
Y -96 ←————→ -92 140 ←————→ 146

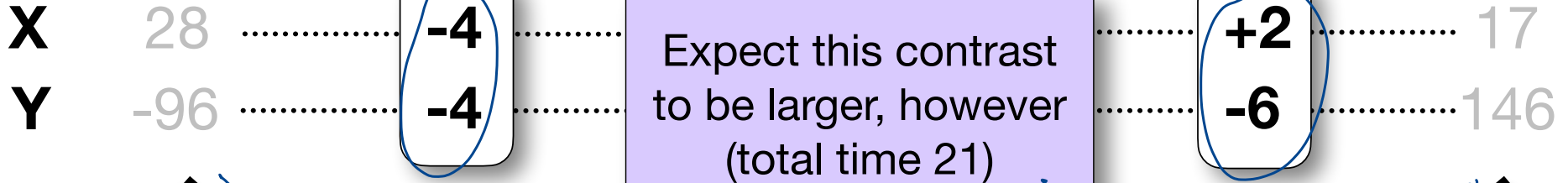


Contrasts (left minus right)



A matter of scale

These contrasts are on the same scale because both have been evolving for the same time (2)



$$\frac{t_i k_j}{t_i + k_j}$$

Why 21 and not 20?



Rescaled Contrasts

Asterisks indicate that the raw contrasts have been scaled by dividing by the square root of the path length.

	X	Y	Variance	X*	Y*
A-B	-4	-4	2	$-4/1.41 = -2.83$	$-4/1.41 = -2.83$
C-D	2	-6	2	$2/1.41 = 1.41$	$-6/1.41 = -4.24$
E-F	12	-237	21	$12/4.58 = 2.62$	$-237/4.58 = -51.7$

These path lengths are proportional to the variance, and thus they are all that are needed to place the 3 contrasts for a given trait on the same scale.

Correlation of Contrasts

	X^*	Y^*
A-B	-2.83	-2.83
C-D	1.41	-4.24
E-F	2.62	-51.7

$$\begin{aligned} \text{Var}(X^*) &= 8.19 \\ \text{Var}(Y^*) &= 773.79 \\ \text{Cov}(X^*, Y^*) &= -54.96 \\ \text{Correlation} &= \mathbf{-0.69} \end{aligned}$$

$$t = -0.95417, \text{ df} = 1$$

$$P = 0.5149$$

not signif.

Correlation of the raw X and Y trait values was -0.95.
Note that the sample size is now 3 rather than 4.

```
cor.test(c(-2.83, 1.41, 2.62), c(-2.83, -4.24, -51.7), alternative="two.sided", method="pearson")
```