# Solutions to Even-Numbered Exercises

to accompany

***An Introduction to Population Genetics: Theory and Applications***

Rasmus Nielsen • Montgomery Slatkin

## CHAPTER 1

1.2 The expected homozygosity, given allele frequencies $f_C = 0.556$ and $f_T = 0.444$, is $1 - 2f_Cf_T = 0.506$. In contrast, the observed homozygosity in the sample is $f_{CC} + f_{TT} = 0.823$. The inbreeding coefficient for this population is

$$F = \frac{2f_Cf_T - f_{CT}}{2f_Cf_T} = 0.639$$

1.4 If 0.02% of the population suffers from a disease caused by a recessive allele *B* that is in HWE with the normal allele *A*, this implies that $f_B = \sqrt{f_{BB}} = \sqrt{0.0002} = 0.0141$. The fraction of the population that carries allele *B* in a heterozygous state is $2f_B(1 - f_B) = 0.0279$, or 2.79%.

1.6 The expected homozygosity is $f_A^2 + f_C^2 + f_T^2 = 0.344$. The expected heterozygosity is 1 – 0.344 = 0.656.

1.8 Let $f_{1a}$ denote the frequency of type *a* at the locus from Exercise 1.1, and let $f_{2a}$ denote the frequency of type *a* at the locus from Exercise 1.5. Assuming HWE, the exact match probability is $\left(2f_{1T}f_{1C}\right) \times \left(2f_{2A}f_{2C}\right) = 0.0970$, or about 9.7%. Using the observed genotype frequencies, the match probability is instead $f_{1CT} \times f_{2AC} = 0.0255$, or about 2.55%.

## CHAPTER 2

2.2 Using Equation 2.3 we find the allele frequency in the next generation to be $(0.1) \times (1 - 10^{-6}) + 0.9 \times 5 \times 10^{-6} = 0.100004$.

2.4 The expected number of substitutions, E[$d$], in the gene is:

$$\begin{aligned} \mathrm{E}[d] &= 2\mu t \\ &= 2\left(\frac{10^{-9}\ \text{mutations}}{\text{generation} \times \text{bp}} \times 800\ \text{bp}\right)\left(6 \times 10^{6}\ \text{years} \times \frac{\text{generation}}{20\ \text{years}}\right) \\ &= 0.48\ \text{mutations} \end{aligned}$$

2.6 As the distribution of the number of offspring is the same for all individuals, and as there are $2N$ individuals producing $2N$ offspring, the expected number of offspring of each individual in the next generation is 1.

2.8 The probability that an offspring chooses a parent without the mutation is $1 - p$. For the mutation to be lost, all $2N$ offspring in the next generation must choose a parent without the mutation. These events are independent, so Pr(*the mutation is lost in one generation*) = $(1 - p)^{2N}$.

## CHAPTER 3

3.2 The observed frequency of pairwise differences per site in the sample is $\pi = 21/10{,}000 = 0.0021$. This leads us to estimate that

$$\hat{N} = \frac{\pi}{4\mu} = \frac{0.0021}{4 \times 10^{-9}} = 525{,}000$$

3.4 Using Equation 3.11, we find $1 - \dfrac{4 \times 20{,}000 \times 10^{-5}}{1 + 4 \times 20{,}000 \times 10^{-5}} = 0.556$

3.6 For the data in Exercise 3.5, Watterson's estimate of $\theta$ is

$$\hat{\theta}_W = \frac{S}{\sum_{k=1}^{3} \frac{1}{k}} = \frac{5}{11/6} = 2.727$$

Tajima's estimate of $\theta$ is $\hat{\theta}_T = \pi = 3$. The numbers could also be reported by standardizing with the number of sites, in which case the estimates would be 0.160 and 0.176 for Watterson's and Tajima's estimators, respectively.

3.8 In a sample of four sequences, the expected proportion of sites that are segregating as singletons in the folded SFS is

$$\frac{1 + 1/3}{1 + 1/2 + 1/3} = 0.727$$

In contrast, we observe that 3/5 = 0.6 of the segregating sites are singletons, slightly fewer than expected under the standard neutral coalescence model.

3.10 In a coalescent tree of sample size $n = 2$, every branch is an external branch, and the expected total branch length is 2. Therefore, the total length of external lineages is 2 for this sample size. For the purpose of induction, suppose that the same is true for a larger sample size $n - 1$. In an $n$-leaf coalescent tree, the expected branch length during the time with $n$ lineages is $2/(n - 1)$, and these are

all external branches. However, exactly $n - 2$ of the tree's external branches extend farther back in time than this point. Each of these external branches is also an external branch of the $(n-1)$-leaf coalescent tree that we get by restricting our study to the time before the first coalescence event occurred. The remaining external branch of the $(n-1)$-leaf tree is ancestral to the two lineages that coalesced most recently in the $n$-leaf tree, so it is actually an internal branch of the $n$-leaf tree. By exchangeability of lineages and the hypothesis of total external branch length being equal to 2, each external branch of the $(n-1)$-leaf tree has expected length $2/(n-1)$. This implies that the $n$-leaf coalescent tree has expected external branch length of $(n-2) \times 2/(n-1) + 2/(n-1) = 2$, which shows that the inductive hypothesis is true for all $n$.

## CHAPTER 4

4.2 a. The allele frequency of *A* in the third population is 0.54, so the average allele frequency in the three populations is $f_A = \frac{1}{3}(f_{A1} + f_{A2} + f_{A3}) = 0.472$.

b. Therefore,

$$H_T = 2f_A(1 - f_A) = 0.498$$
$$H_S = \frac{2}{3}\left(f_{A1}(1 - f_{A1}) + f_{A2}(1 - f_{A2}) + f_{A3}(1 - f_{A3})\right) = 0.488$$
$$F_{ST} = \frac{H_T - H_S}{H_T} = \frac{0.498 - 0.488}{0.498} = 0.020$$

4.4 a. In coalescent-scaled time, $E_S[t] = 2$. This corresponds to $2 \times 10^4$ generations.

b. $M = 2Nm = 10^4 \times 10^{-4} = 1$, so $E_D[t] = 2 + 1/(2M) = 2.5$. This is $2.5 \times 10^4$ generations.

c. Finally, $F_{ST} = 1/(1 + 8M) = 0.11$.

4.6 First rearrange $F_{ST} = 1/(1 + 8M)$ to get

$$M = \frac{1}{8}\left(\frac{1}{F_{ST}} - 1\right)$$

Furthermore, $M = 2Nm$, so we get

$$m = \frac{1}{8 \times 2N}\left(\frac{1}{F_{ST}} - 1\right)$$
$$= \frac{1}{8 \times 10^4}\left(\frac{1}{0.016} - 1\right)$$
$$= 7.7 \times 10^4$$

4.8 We follow the arguments for the case of two populations, but use values of $E_S[t]$ and $E_D[t]$ for an island model with *d* demes.

We get that $H_S = d\theta/k$. For the case of $H_T$, note that the probability that two samples are from the same population is $1/d$, so we have that

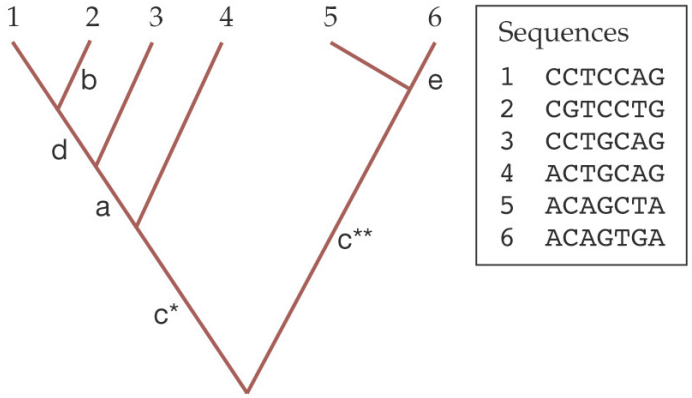
$$H_T = \frac{1}{d}\frac{d\theta}{k} + \frac{d-1}{d}\frac{\theta[1/(2M)+d]}{k} = \frac{\theta}{k}\left(d + \frac{d-1}{2dM}\right)$$

Plugging into the definition of $F_{ST}$ gives:

$$\begin{aligned} F_{ST} &= \frac{H_T - H_S}{H_T} \\ &= \frac{(d-1)/(2dM)}{d + (d-1)/(2dM)} \\ &= \frac{(d-1)/d}{2Md + (d-1)/d} \end{aligned}$$

## CHAPTER 5

5.2




Position 1: Mutation A ⇒ C labeled ‘a’ in the figure

Position 2: Mutation C ⇒ G labeled ‘b’ in the figure

Position 3: Mutation A ⇒ T (or T ⇒ A) labeled ‘c*’ (or ‘c**’) in the figure

Position 4: Mutation G ⇒ C labeled ‘d’ in the figure

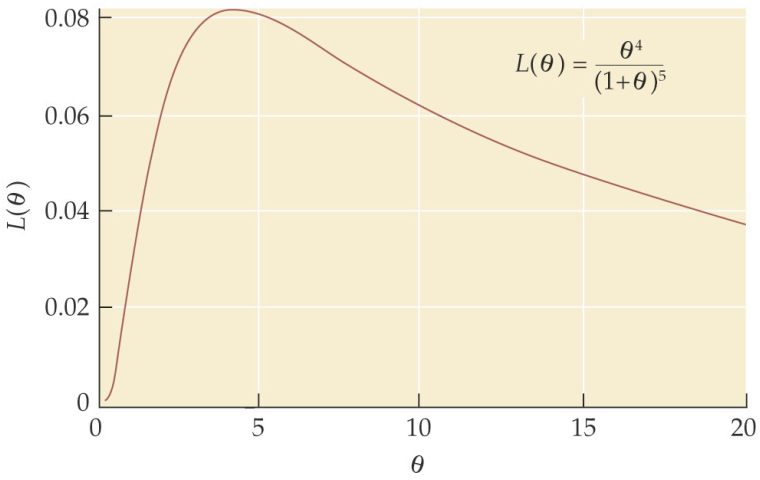
Position 5: Mutation C ⇒ T labeled ‘e’ in the figure

Position 6: Not marked, because the mutation in position 6 is not compatible with this tree under infinite sites

Position 7: Mutation A ⇒ G (or G ⇒ A) labeled ‘c*’ (or ‘c**) in the figure (same as position 3)

5.4 The likelihood function $L(\theta)$ can be computed as follows:

$$L(\theta) = \frac{\theta^4}{(1+\theta)^5}$$



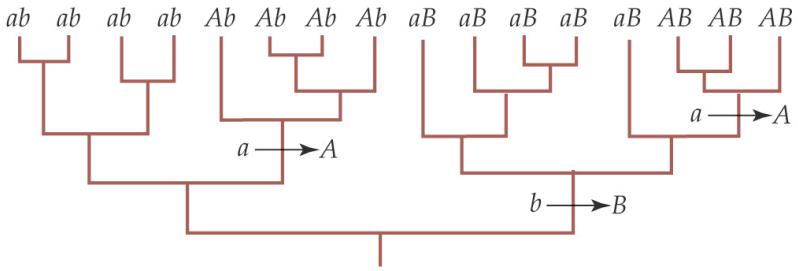

As seen in this plot, $L(\theta)$ is maximized around $\theta = 4$. This agrees with Watterson's estimate $\hat{\theta}_w$ for sample size $n = 2$:

$$\hat{\theta}_w = \frac{4}{\sum_{j=1}^{2-1} \frac{1}{j}} = 4$$

## CHAPTER 6

6.2 Two haplotypes out of four can be missing only if one of the allele frequencies is 0 or if allele frequencies at the two loci are equal.

6.4 The tree below shows the answer to part a. The haplotype counts are *AB* (3), *Ab* (4), *aB* (5), *ab* (4). $f_A = 7/16$, $f_B = 1/2$, $D = 3/16 - (7/16)(1/2) = -1/32$, $D' = 0.215$.




6.6 a. The allele frequencies are $f_A = 0.3$, $f_a = 0.7$, $f_B = 0.4$, $f_b = 0.6$. The sample size is 1000. The $\chi^2$ statistic is computed by summing (observed–expected)$^2$/expected over the four haplotype categories:

$$\chi^2 = \frac{(30-120)^2}{120} + \frac{(370-280)^2}{280} + \frac{(270-180)^2}{180} + \frac{(330-420)^2}{420} = 160.7$$

Therefore, there is significant LD at the 1% level.

b. $D = 0.03 - (0.3 \times 0.4) = -0.09$

c. In Box 6.4, $\chi^2 = nD^2/(f_A f_B f_a f_b)$. The question is, given $n$ and the allele frequencies, what is the largest $|D|$ can be so that $\chi^2 \leq 6.636$. The largest absolute value of $D$ is $\sqrt{6.636 f_A f_B f_a f_b / n} = \sqrt{6.636 \times 0.3 \times 0.4 \times 0.6 \times 0.7 / 1000} = 0.0183$. The sign does not matter.

d. The question is how long it takes for $D$ to decrease from its initial value of 0.03–(0.3 × 0.4) = –0.09 to –0.0183. From Equation 6.8, $D(t) = D(0)(1 - c)^t$.

Solving for $t$: $t = \frac{\ln[D(t)/D(0)]}{\ln(1-c)} = \frac{\ln(0.0183/0.09)}{\ln(0.999)} \approx 1592$ generations.

6.8 a. $1/c$ =1000 generations. The mean of an exponential distribution is $1/c$.

b. The population size does not matter because the two lineages are on a single chromosome. The population size affects the probability of coalescence.

c. They are joined on the same chromosome if the two chromosomes coalesce, which takes an average of $2N$ generations. Therefore, the waiting time is 200 generations for the smaller population and 2,000,000 generations for the larger population.

d. The rate of coalescence depends only on the population size.

e. The average time the two lineages remain on separate chromosomes is given by the ratio of the average waiting times, $c/[c + 1/(2N)] = 2Nc/(1 + 2Nc)$. For $N$ = 100, the ratio is 0.2/1.2 = 0.167. For $N$ = 1,000,000, the ratio is 2000/2001 ≈ 0.9995.

## CHAPTER 7

7.2 In a population of size 10,000, the initial frequency of the mutant is 1/20,000, so that is $f_A(0)$. From Box 7.1, $t = 1204$ generations If $N = 100{,}000$, the initial frequency is 1/(200,000), so $t = 1433$ generations.

7.4 a. Because the allele is a recessive lethal, the equilibrium frequency is given by Equation 7.5 with $s = 1$. Therefore $0.02 = \sqrt{\mu}$, which implies that $\mu = 4 \times 10^{-4}$. That mutation rate is at least two orders of magnitude larger than mutation rates at other loci, which suggests that CF does not result from alleles under mutation–selection balance.

b. At equilibrium, Equation 7.3 implies $0.98 = 1/(s_{AA}+1)$ or $s_{AA} = 0.02/0.98 \approx 0.0202$. It does not require much of a heterozygote advantage, but no one has found what advantage heterozygous carriers of CF alleles have, if indeed they have one.

7.6 At HWE, the genotype frequencies are $f_{BB} = 1/16$, $f_{Bb} = 6/16$ and $f_{bb} = 9/16$. Among adults, $f'_{BB} = (1/16)(1/5)/\bar{v}$, $f'_{Bb} = (6/16)(1/6)/\bar{v}$, and $f'_{bb} = (9/16)(1/10)/\bar{v}$, where $\bar{v} = (1/16)(1/5) + (6/16)(1/6) + (9/16)(1/10) = 21/160 = 0.13125$. Therefore, $f'_B = f'_{BB} + f'_{Bb}/2 = 1/3 = 0.3333$. Thus, $f_B = 0.3333$ in the seeds the next generation, whether or not the adults mate randomly, as long as there is no effect on fertility of the *B/b* locus.

7.8 The initial frequency is 1/20,000. The viability of the heterozygotes for the darker allele is 80%–90% of the viability of the individuals homozygous for the light allele, so $s = 0.1 - 0.2$. From Box 7.1, $t = 137$ generations if $s = 0.1$ and 65 generations if $s = 0.2$.

7.10 a. Here is a table describing the fertility of each family:

| Father | Mother | Fertility |
|---|---|---|
| *AA* | *AA* | 1 |
| *AA* | *Aa* | ½ |
| *AA* | *aa* | 1 |
| *Aa* | *AA* | ½ |
| *Aa* | *Aa* | ¼ |
| *Aa* | *aa* | ½ |
| *aa* | *AA* | 1 |
| *aa* | *Aa* | ½ |
| aa | aa | 1 |

b. If *A* is rare, *Aa* × *aa* families will be less fertile than *aa* × *aa* families. Therefore, *A* will tend to decrease in frequency. If *a* is rare, *Aa* × *aa* families will be less fertile than *AA* × *AA* families, so *a* will tend to decrease in frequency.

7.12 a. With 5 *S* alleles, there are ten possible heterozyotes ($S_1S_2$, $S_1S_3$, $S_1S_4$, $S_1S_5$, $S_2S_3$, $S_2S_4$, $S_2S_5$, $S_3S_4$, $S_3S_5$, and $S_4S_5$). By symmetry, they have to have equal frequencies, 0.1.

b. Let $S_6$ be the new mutant, which is found initially in one plant, which has the genotype $S_1S_6$. Half of the pollen produced by this plant will carry $S_1$ and will not be able to fertilize 4/10 of the other plants, those that have genotypes $S_1S_2$, $S_1S_3$, $S_1S_4$, and $S_1S_5$. Half of the pollen produced by this plant will be able to fertilize every other plant. Contrast that with any other plant, say, an $S_1S_2$ plant. The $S_1$-bearing pollen cannot fertilize 4/10 of the other plants, and the $S_2$ pollen also cannot fertilize 4/10 of the other plants. Therefore, the mutant plant will be able to fertilize 2/10 more plants than any other (1/2 of 4/10).

## CHAPTER 8

8.2 $s$ is positive in this case.

a. $N = 10{,}000$, $r = 8.8 \times 10^{-9}$; $N = 1000$, $r = 9.0 \times 10^{-9}$; $r = 2.6 \times 10^{-9}$.

b. $N = 10{,}000$, $r/\mu = 40$; $r/\mu = 4.1$; $r/\mu = 1.2$.

8.4 The formula is $r = d/(2LT)$. $L = 1000$, $d = 4$ and $T = 6 \times 10^6$ years, so, $r = 0.333 \times 10^{-9}$ per nonsynonymous site per year.

8.6 The total substitution rate, $r$, is the sum of the rates for neutral alleles $(1 - \alpha)\mu$ and for the slightly deleterious alleles, $(\alpha/3)2N\mu u(s,N)$, where $N = 100$, $s = 0.001$, and $\mu = 2.2 \times 10^{-9}$. That is $0.8 = (1 - \alpha)2.2 + (\alpha/3)1.8$ which implies that $\alpha \approx 0.88$. This problem illustrates that $\alpha$ computed from Equation 8.4 is the minimum fraction of deleterious nonsynonymous mutations consistent with the observed rate.

8.8 $t = 1n[(\delta - f_B)/(1 - f_B)]/1n(1 - c)$ where $f_B = 0.086$, $c = 0.023$ and $\delta = 47/54 = 0.087$. Therefore $t \approx 6.5$ generations ago. Assuming 25 years per generation, the causative mutation probability arose within the last two hundred years or so.

8.10 The model is equivalent to one with ten islands, each of size $N/10$.

a. The effective migration rate is the probability that recombination will move the neutral locus on one genetic background (say $S_1$) to any other genetic background ($S_2$, . . . $S_{10}$). Because there are no homozygotes, every copy of $S_1$ is in a heterozygote, so every recombination event will move the neutral locus to another genetic background. Therefore, the effective migration rate is $c$, the recombination rate.

b. Given that one neutral copy is on a particular background (say $S_1$), the probability that the other is on the same background is 1/10, and the probability that the other is on a different background is 9/10. Therefore, using the result on p. 69 of Chapter 4, $H_T = \left[\frac{1}{10}10 + \frac{9}{10}\left(10 + \frac{1}{2M}\right)\right]\frac{\theta}{k} = \left[1 + \frac{9}{10}\frac{1}{2M}\right]\frac{\theta}{k}$.

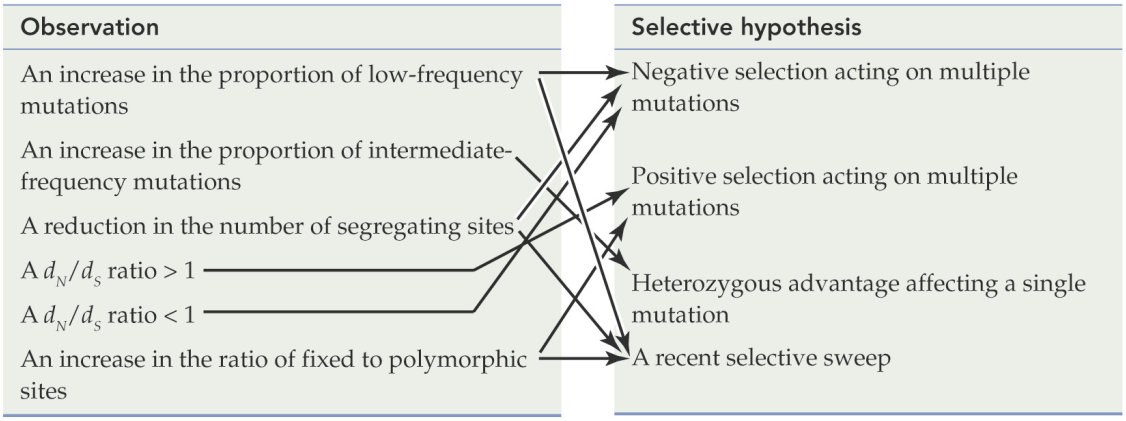
In the island model, $M = 2Nm$, so in this model, $M = 2Nc/10$, and the result is

$$H_T = \left[1 + \frac{9}{4Nc}\right]\frac{\theta}{k},$$

c. In the above expression, the number of $S$ alleles is present only in the numerator of the fraction, $9/(4Nc)$. As the number of $S$ alleles increases, $H_T$ will increase, also.

## CHAPTER 9

9.2




9.4 There are 8 segregating sites, so $\hat{\theta}_w = \frac{8}{\sum_{j=1}^{5-1} \frac{1}{j}} = 3.84$. Similarly, we find

$$\hat{\theta}_T = \frac{2\times3+1\times4+1\times4+2\times3+1\times4+1\times4+2\times3+2\times3}{\binom{5}{2}} = 4$$

As Watterson's estimate is smaller than Tajima's estimate, Tajima's $D$ would be positive. Positive values of Tajima's $D$ are generally compatible with negative selection, although in this case, the value of Tajima's $D$ would not be negative enough to allow us to reject the neutral null hypothesis with statistical confidence.

## CHAPTER 10

10.2 $\frac{m'}{Nm}+\frac{f'}{Nf}=\frac{m+\delta m}{Nm}+\frac{1-m-2\delta m}{N(1-m)}=\frac{m}{Nm}+\frac{1-m}{N(1-m)}+\frac{\delta m}{N}\left(\frac{1}{m}-\frac{2}{1-m}\right)$. The coefficient of $\delta m$ will be 0 when $m = 1/3$.

10.4 The problem tells you that $\hat{f}=0.5$ and that the decrease in mortality when a mouse is in a hole is 0.2. If two mice display, then each gains a 0.1 increase on average. That is *d* in Table 10.1. If one displays and the other fights, the gain to the other is 0.2, assuming no cost to displaying. That is *c* in Table 10.2. You are asked to find *b* by solving Equation 10.5, which gives $b = 0.1$. You would conclude that there is a 10% increase in the risk of mortality to each individual if they both fight.

10.6 The observations are consistent with the theory of kin selection, which predicts that altruistic behavior should be preferentially directed to close relatives. They are not consistent with the theory of group selection. Warning calls directed to non-relatives would help promote the survival of the group.

10.8 With $f_A(0) = 0.1$, $f_A^* = 0.9$ and $s = 0.01$, $t \approx 439$ generations.

10.10 From Equation 10.18, the advantage to *B* is approximately

$$\left[1+\frac{s(f_a-f_a')}{1-sf_a}\right]^2$$

With $m = 0.1$ and $s = 0.5$, $f_a = m/s = 1/5$. From Equation 10.14, $f_A' = 1/9$. Therefore, the offspring carrying a *B* has a viability 4/81 larger than that of the offspring of a *bb* individual.

## CHAPTER 11

11.2 a. Using the formula $Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$, where $x_i$ is the weight of the father {650, 670, 700, 630, 680 g} and $y_i$ is the weight of the offspring {687, 618, 618, 600, 717 g}, Cov($x$,$y$) = 186 g$^2$. The additive variance is twice the covariance when only one parent is used. Therefore, $V_A$ = 372 g$^2$.

b. The variance among the fathers is 584, so $h^2$ – 372/584 = 0.64.

11.4 For males, $\alpha = 0.55$ and $\delta = 0.35$. For females, $\alpha = 0.3$ and $\delta = 0.2$.

11.6 $\bar{x} = (0.4)^2(7) + 2(0.4)(0.6)(8) + 0.6^2(9) = 8.2$. Note that $\alpha = 1$ and $\delta = 0$. Therefore, $V_A = 2(0.4)(0.6) = 0.48$ and $V_x = V_A + V_E = 1.48$.

11.8 From Equation 11.13, $V_A = 2f_A f_a\left[\alpha - (f_A - f_a)\delta\right]^2$. Therefore, $V_A = 0$ when $\alpha = (f_A - f_a)\delta$ or $f_A = (\alpha + \delta)/(2\delta)$. Because $\delta > \alpha$, $0 < f_A < 1$.