

Probability Theory and Mathematical Modelling: A Primer

Readings: Chp. 1 (Hamilton)
algebra review sheet (download from course website)

If you struggle with this material, read the following sections in *Introduction to Population Genetics* by Richard Halliburton (on course reserve in Morisset):
Appendix A, Box 2.2, and section 1.4

Learning objectives

By the end of this section, you should:

- Know the rules of basic probability
- Be able to recognize whether outcome probabilities are mutually exclusive, overlapping, or conditional, and how to solve questions regarding probability
- Appreciate the utility of, and interpret output from basic mathematical models (e.g., graphical analysis, finding equilibria, assessing stability)

MATH!

- In population genetics, we use math (probability, modelling and simulations) to understand the genetic consequences of different evolutionary processes.
- With an understanding of basic probability, we can understand the rules of inheritance.
- With rudimentary mathematical modeling and simulation studies, we can predict and understand how populations will evolve.

Probability basics

- Probability is a measure of the likelihood of some event, “X”, occurring.
- We’ll symbolize it ‘Pr’ (because P will be used for genotype/gamete frequency, p for allele frequency, and ‘P-value’ for the probability from statistical tests)
- $0 \leq \text{Pr} \leq 1$ **ALWAYS**

Probability basics

- What is the probability of getting “tails” from a single coin flip?



- Rolling a 6?



- Rolling a 12?



Probability basics

- A probability trial is any sort of occurrence (e.g., the birth of a child, a pair of dice are rolled, selecting alleles from a gene pool)
- The probability of X is the # of times X occurs out of a large number of trials: $\text{Pr}(X) = \# \text{ times outcome } X \text{ occurred} / N$
(i.e. the probability of X is its relative frequency over the long run)
- Alternatively, $\text{Pr}(X)$ is:
of outcomes that include X / # of all possible outcomes
e.g., $\text{Pr}(\text{sum of } 7 \text{ on two dice})$
= # ways you can get 7 / total # of combinations of 2 dice
= 6 / 36

Question:

What is the probability of a heterozygote (i.e. Aa) passing on allele A to its offspring?

Rule 1

- For the **complete set** of all possible outcomes, the probabilities must **sum to one**:

$$\sum_{i=1}^{\text{\#of outcomes}} \Pr(X = x_i) = 1$$

- In other words, there is a 100% chance of *something* happening.

The complement

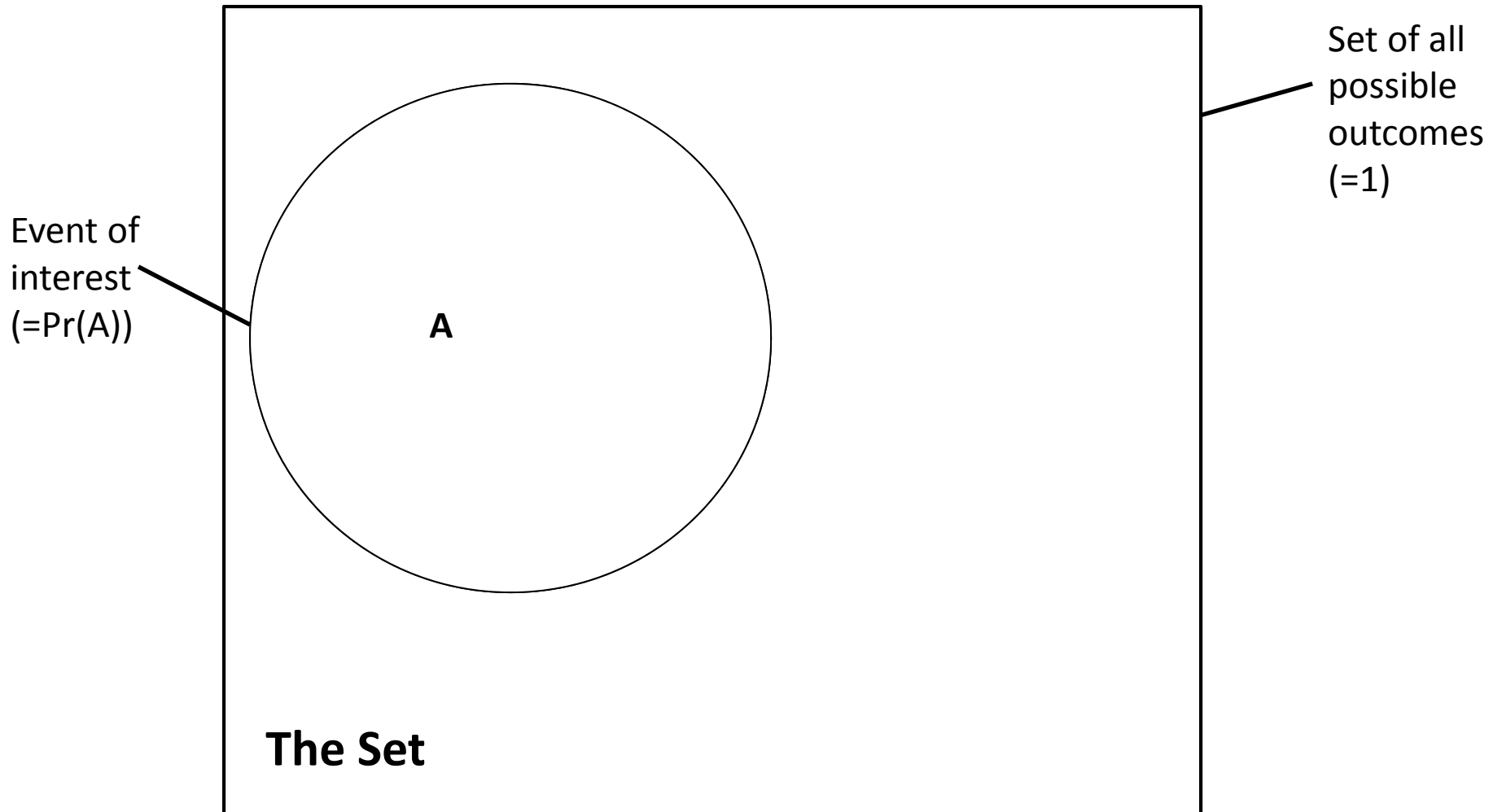
- The complement is the set of possible outcomes not considered in the present case

e.g. The complement of A (A^C) is the probability of “A” not happening.

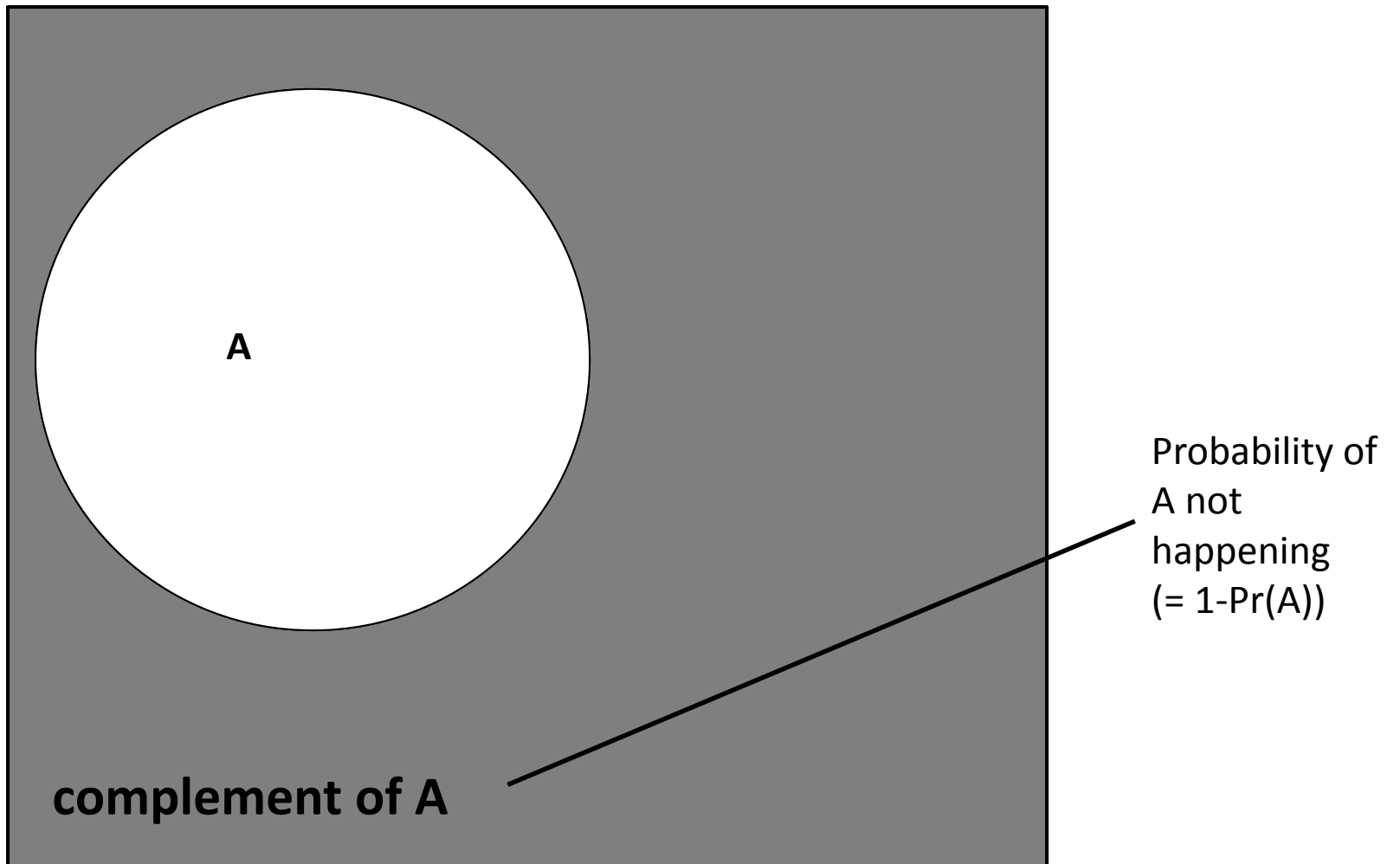
If $\Pr(X=A) = 0.7$, then the compliment of A is

$$\Pr(X \neq A) = 1 - 0.7 = 0.3$$

The complement



The complement



Rules 2: The Complement

- The probability of an outcome plus the probability of its complement must sum to one (this is used often in population genetics):

$$\Pr(X = A) + \Pr(X = A^c) = 1$$

Example:

Say we know that the probability that a baby is going to be born on a Monday is ($A = 1/7$). The complement A^c is the probability that it is born on any other day of the week. Therefore, the probability that the baby is *not* born on a Monday is:

$$\Pr(X=A^c) = 1 - \Pr(X=A) = 1 - 1/7 = 6/7$$

The complement and population genetics

- Many of you are probably already familiar with the p 's and q 's notation for representing **allele** frequencies at a single locus in a population. When there are only two alleles, p denotes the frequency of one (say A or A_1) and q the frequency of the other (say a or A_2)
- If there are **only two alleles at a locus**, then p and q are each other's complement:
$$q = 1 - p \text{ and } p = 1 - q \text{ because } p + q = 1$$

Exercise

- Say there is a locus with three alleles, A_1 , A_2 , and A_3 .
- Allele frequency of A_1 is $p_1 = 0.3$
- What is the probability of taking a random allele from this population, and finding it is not allele A_1 ?

Asides:

- don't confuse $\Pr(X)$ and p in your notation.
- An allele frequency can be thought of as the probability of choosing an allele of that type in a random draw from the population (gene pool): e.g. $p = \Pr(X=A_1)$

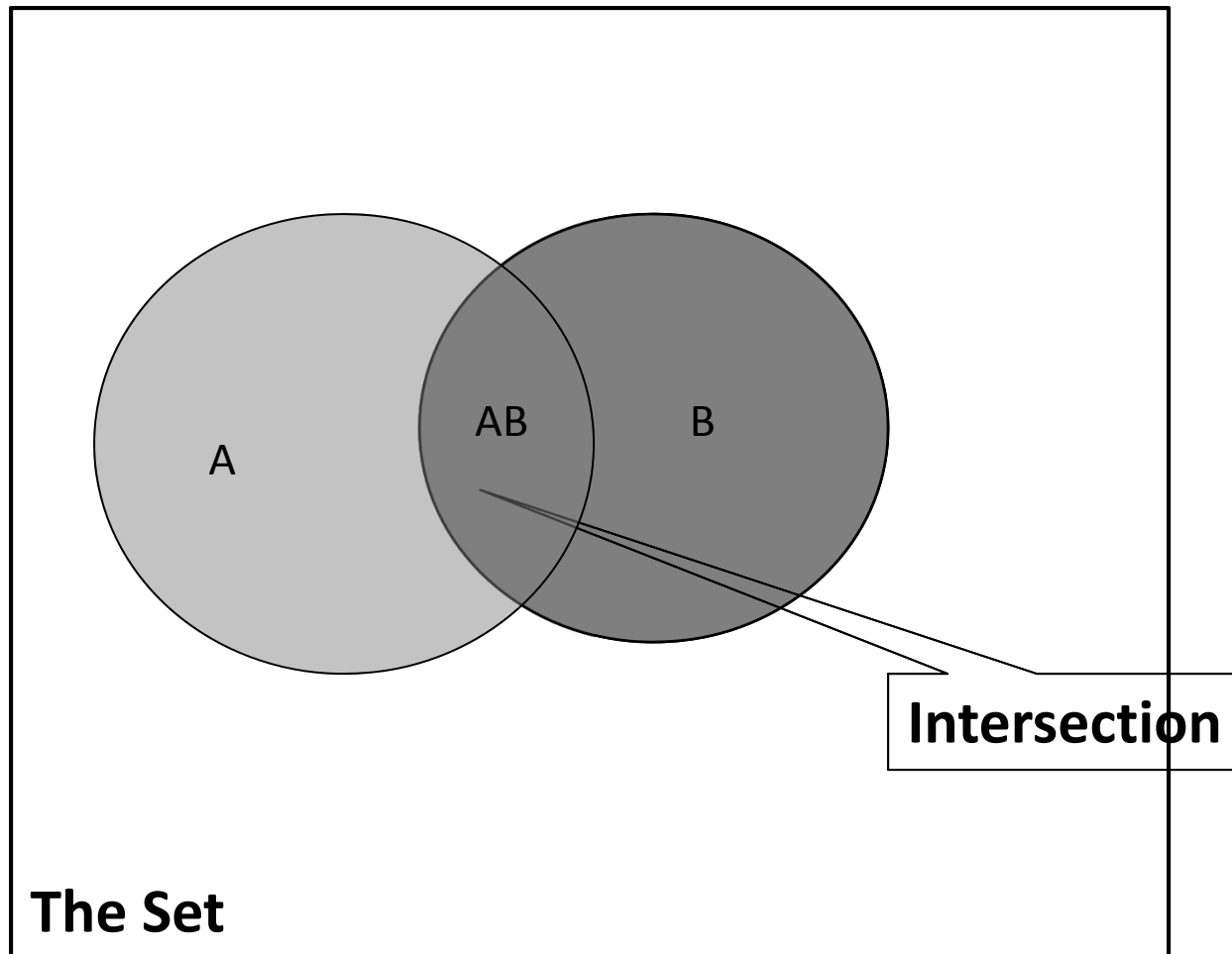
Rule 3a: The Law of Addition

The probability that **either** (or both) of two events will occur is the sum of their individual probabilities minus the probability that both will occur simultaneously:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

Why this latter term?

...to avoid counting the intersection twice



Example

- What is the probability that an offspring of a double heterozygote ($AaBb$) receives an A or a B allele from this parent?

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$
$$=$$

Note: $\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B)$. This is Rule 4 which we'll cover shortly.

Or count the outcomes:

Gamete haplotype
at the 'B' locus

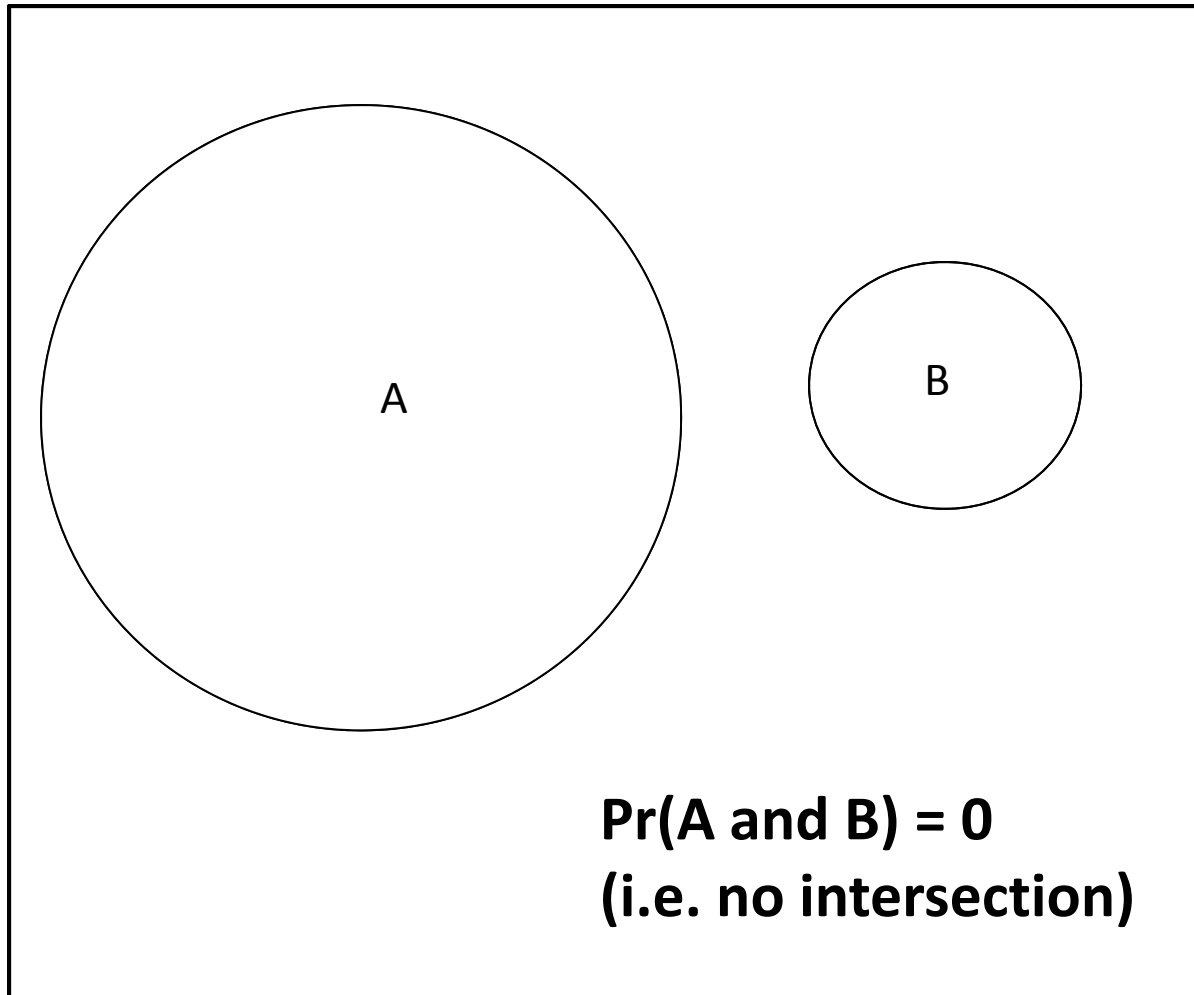
Gamete haplotype at the 'A' locus

	A	a
B	AB	aB
b	Ab	ab

Three of four gamete types carry and A or a B allele = $\frac{3}{4} = 0.75$

(this assumes Mendel's second law: i.e. segregation of alleles the A locus is independent of that at the B locus – we'll talk about this later)

Consider mutually exclusive outcomes
(events are mutually exclusive if it is impossible
for them to both occur.)



Rule 3b: The Law of Addition

- For **mutually exclusive outcomes** $\Pr(A \text{ and } B) = 0$ (by definition), so if A and B are mutually exclusive:

$$\begin{aligned}\Pr(A \text{ or } B) &= \Pr(A) + \Pr(B) - \Pr(A \text{ and } B) \\ &= \Pr(A) + \Pr(B) - 0 \\ &= \Pr(A) + \Pr(B)\end{aligned}$$

Example

What is the probability of an offspring of an Aa heterozygote parent receiving an A or an a allele from this parent?

These are mutually exclusive outcomes (a given parent can't pass on both an A and an a allele to a particular offspring), so:

$$\Pr(A \text{ or } a) = \Pr(A) + \Pr(a)$$
$$=$$

Independence

- Two events are independent if the occurrence of one of the events provides no information about whether or not the other event will occur
- In other words, the events have no influence on one another

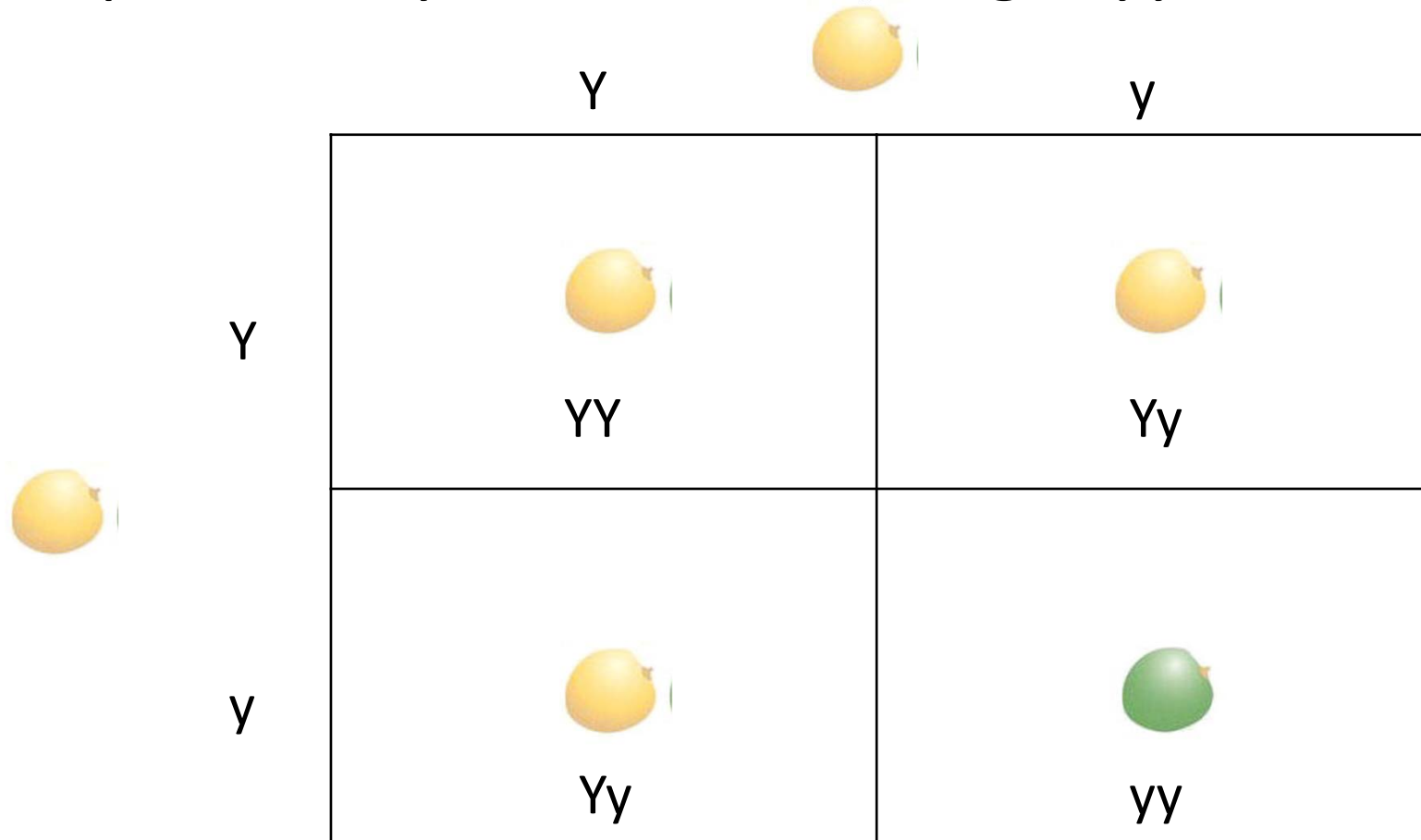
Rule 4: Product rule for **independent events**

- If two events are **independent**, then the probability that they both occur is equal to the product of the probabilities of the two individual events:

$$\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B) = \Pr(A)\Pr(B)$$

Example







- If two parents of genotype Yy mate, what is the probability that the offspring is yy ?



The allele inherited from your father is independent of the allele you inherit from your mother, so

$$\Pr(yy) = \Pr(y \text{ from mom}) \times \Pr(y \text{ from dad}) = 0.5 \times 0.5 = 0.25.$$

Or count outcomes:

		Y 	y
Y 	 YY	 Yy	
y	 Yy	 yy	

In class Problem

- Say two double heterozygotes ($AaBb$) produce an offspring.
- What is the probability that the offspring will receive an A or B allele?
- Be careful – there are two things going on here: the alleles received from parent 1 and alleles received from parent 2

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

Not mutually
exclusive

where

$\Pr(A)$ is the prob of receiving the A allele from either parent

$$= \Pr(A \text{ from } P1 \text{ or } P2)$$

$$= \Pr(A \text{ from } P1) + \Pr(A \text{ from } P2) - \Pr(A \text{ from } P1 \text{ and } P2)$$

=

Not mutually
exclusive

And, $\Pr(B)$ is the prob of receiving the B allele from either parent

$$= \Pr(B \text{ from } P1) + \Pr(B \text{ from } P2) - \Pr(B \text{ from } P1 \text{ and } P2)$$

=

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

where

$\Pr(A)$ is the prob of receiving the A allele from either parent

=

$\Pr(B)$ is the prob of receiving the B allele from either parent

=

So:

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

=

Assume A and B are independent so
 $\Pr(A \text{ and } B) = \Pr(A) \times \Pr(B)$

Or identify all possible outcomes and then count those that meet the desired criteria:

		Gamete haplotype from parent 1			
		AB	Ab	aB	ab
Gamete haplotype from parent 2	AB	AABB	AABb	AaBB	AaBb
	Ab	AABb	Aabb	AaBb	Aabb
	aB	AaBB	AaBb	aaBB	aaBb
	ab	AaBb	Aabb	aaBb	aabb

Conditional Probability

- Given that outcome B has occurred, the probability of observing outcome A is denoted: $\Pr(A|B)$.
- For independent events, observing B gives **no information** about A :

$$\Pr(A|B) = \Pr(A)$$

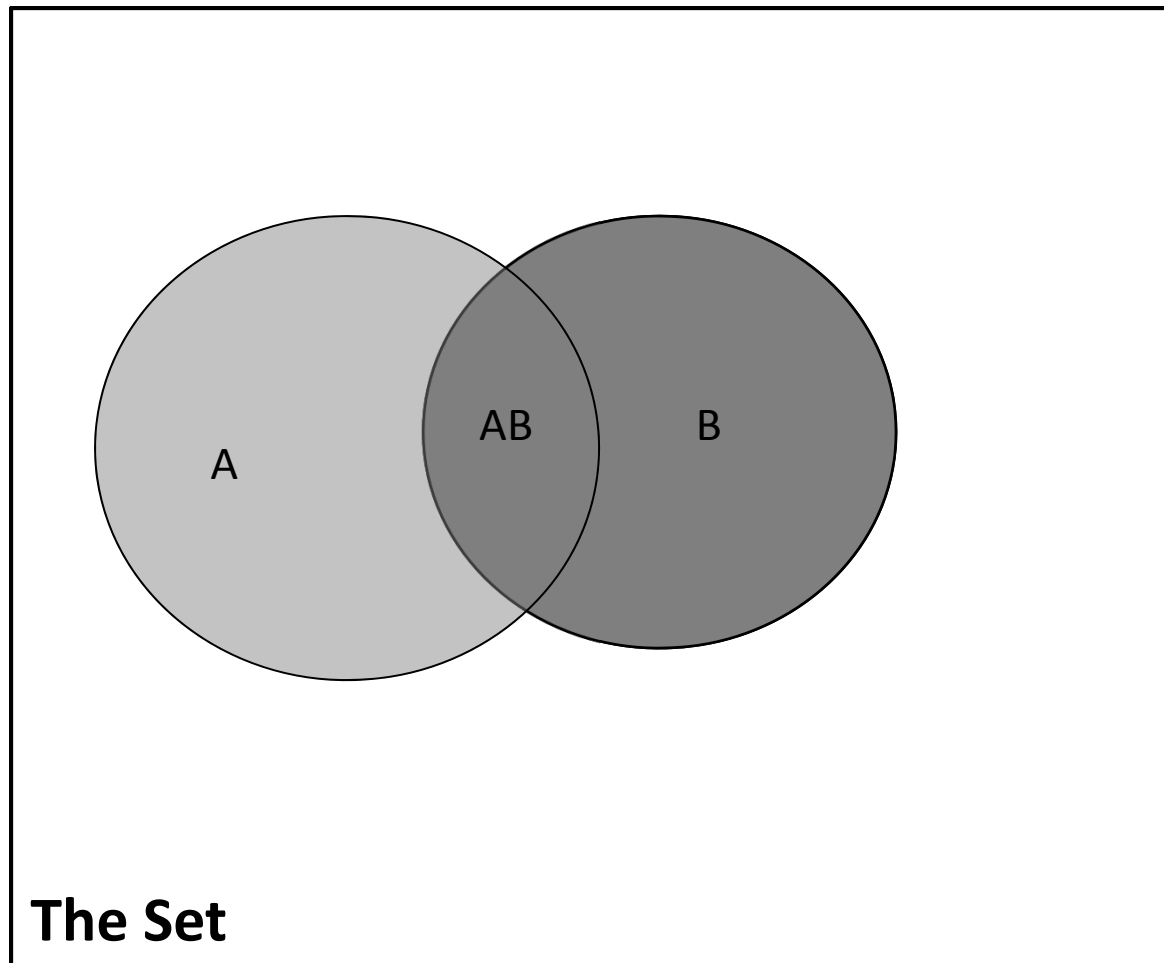
- For example, whether a parent donates an A or an a allele given that they donated a B allele (assuming these loci are unlinked).

Rule 5: Conditional Probability (for events that are **non-independent**)

- If the outcome of B affects the outcome of A, then the probability of A given that B has occurred is:

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

$\Pr(A | B)$ represents the fraction of event B that falls into the union space of A and B:



In-class Exercise

- If the probability of owning a hybrid car is 3%, the probability of voting for the Green Party is 15%, and the probability of both owning a hybrid car and voting Green is 2%, what is the probability that you vote Green given that you own a hybrid car?

$$\Pr(A|B) = \frac{\Pr(A \text{ and } B)}{\Pr(B)}$$

The Joint Probability

- The joint probability is simply the probability that two outcomes are both observed.
- This is the probability of observing B **times** the probability that given the occurrence of B , A also occurs (can you see how this is a rearrangement of the previous?)

$$\Pr(A \text{ and } B) = \Pr(B)\Pr(A|B)$$

$$\Pr(A \text{ and } B) = \Pr(A)\Pr(B|A)$$

to recap...

- Recall that for *independent* outcomes, observing B gives **no information** about A :

$$\Pr(A|B) = \Pr(A)$$

to recap...

- Recall that for *independent* outcomes, observing B gives **no information** about A :

$$\Pr(A|B) = \Pr(A)$$

- Can you see how the joint probability reverts to the familiar **product rule** if A and B are independent?

$$\begin{aligned}\Pr(A \text{ and } B) &= \Pr(B)\Pr(A|B) \\ &= \Pr(B)\Pr(A)\end{aligned}$$

Rule 6: Bayes' Theorem

- Because of the fact that $\Pr(A \text{ and } B)$ can be calculated from either joint probability, we can calculate one conditional probability from the other using Bayes' Theorem:

$$\Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$$

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}$$

Exercise

- Two parents have the genotypes AaBb and AaBB
- What is the probability that they will have a female offspring carrying either the A or b allele (or both)?

Exercise

- What is the probability that they will have a female offspring carrying either the A or b allele (or both)?

$$\Pr(\textit{Girl} \text{ and } (A \text{ or } b)) = \Pr(\textit{Girl}) \times (\Pr(A) + \Pr(b) - \Pr(Ab))$$

=

Part 2: Probability Distributions

- Describes the probability that a random variable takes on various values.
- Two forms of distributions: discrete and continuous

Random Variables

- Random variables are used to help us understand an unknown value about which we may have some information.
- For example, if we flip a coin, we don't know what the outcome will be, but we know that there are only two options.
- We actually have even more information than that, because, we know that both options are equally likely to occur.
- Thus, the number of heads appearing from a single coin toss can be described as the random variable

$$X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$$

Discrete Distributions 1:Bernoulli Trial

- a Bernoulli Trial is a trial that has two possible outcomes, say 'zero' and 'one' or 'heads' and 'tails'. The probability of the random variable X taking on different outcomes is described as:

$$\Pr(X = 1) = p$$

$$\Pr(X = 0) = 1 - p$$

The Binomial Distribution

- Is a generalization of a single Bernoulli trial to n independent trials.
- p is the probability of a 'success' in a single Bernoulli trial – such as a coin flip
- the binomial describes the probability of observing k 'successes' in n trials (say, 5 heads in 10 flips)

Discrete Distributions: The Binomial Distribution

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Expected value (i.e. mean) of X : $E(X) = np$

Variance of X : $Var(X) = np(1-p)$

In class exercise

- If there is a 50% probability of giving birth to a girl ($p = 0.5$), the binomial distribution can describe the probability of observing three girls in a family of five children. Calculate this probability.

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

- In a family of $n=6$ children, how many girls would you expect?

Discrete Distributions: The Poisson Distribution

- Describes the random occurrence of events that are individually rare but for which there are many opportunities for the event to occur (i.e. trials)
- the probability of observing k events in a given space or time period when the expected number of events is μ and when each event occurs independently:

$$\Pr(X = k) = \frac{e^{-\mu} \mu^k}{k!}$$

In class Exercise

- Births in a hospital occur randomly at an average rate of 2 births per hour. What is the probability of observing two or more births in a given hour at the hospital?

$$\Pr(X = k) = \frac{e^{-\mu} \mu^k}{k!}$$

In class Exercise

- Births in a hospital occur randomly at an average rate of 2 births per hour. What is the probability of observing two or more births in a given hour at the hospital?

We want: $\Pr(X \geq 2) = 1 - (\Pr(X = 1) + \Pr(X = 0))$

$$\Pr(X \geq 2) =$$

Mathematical modeling

- How we see the world



Mathematical modeling

- How a mathematician sees the world

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

- Mathematical models are a simplification of reality that can be used to clarify hypotheses/assumptions and make predictions.
- Think of a model as an equation that gives 'expected values' based on a series of observations

- With *dynamic* models, the goal is to describe how a system changes over time.
- We make predictions about what's going to happen in the future based on present conditions.
- E.g. how do allelic frequencies change with mutation?

$$\left(\begin{array}{c} \textit{frequency of A} \\ \textit{next generation} \end{array} \right) = \left(\begin{array}{c} \textit{A's frequency} \\ \textit{now} \end{array} \right) - \left(\begin{array}{c} \textit{proportion of A} \\ \textit{that mutates to a} \end{array} \right)$$

$$p_{t+1} = p_t - \mu p_t$$

- We know something about the present (p_t) and are interested in what will happen in the future (p_{t+1}) given known variables (e.g. mutation rate) and established laws (the equation).

$$p_{t+1} = p_t - \mu p_t$$

- With a given formula, predicting the outcome is a simple matter of plugging in numbers.

$$p_{t+1} = p_t - \mu p_t$$

- With this formula, we can find the value of any of the three variables (p_t, p_{t+1}, μ) using simple arithmetic and information on the other two variables.
- This is known as a *recursion equation* which describes how the variable changes across one time step

- e.g. if $p_t = 1$ and $p_{t+1} = 0.95$, what is μ ?
- Just plug the numbers in and solve for μ .

$$p_{t+1} = p_t - \mu p_t$$

- Many models we'll be working with also have a general solution.
- This means that we do not need to make predictions one time step (e.g. generation) at a time.
- For example, the previous equation's general solution is $p_t = p_0(1 - \mu)^t$, where p_0 is the initial allele frequency, and t is the number of times steps.
- We could make predictions for the 3rd or 100th or any future generation with this equation.

- Equations will get more complicated than this, but to make predictions and calculate variables, it's just a matter of plugging in numbers and solving for the unknowns.

Two flavours of models

Discrete time

- Processes being studied occur in discrete time intervals (e.g., generations, years), like mutation
- Models formulated via recursion equations that predict the value one time step in the future from the current value
- Often harder to solve explicitly and tend to yield more complicated results

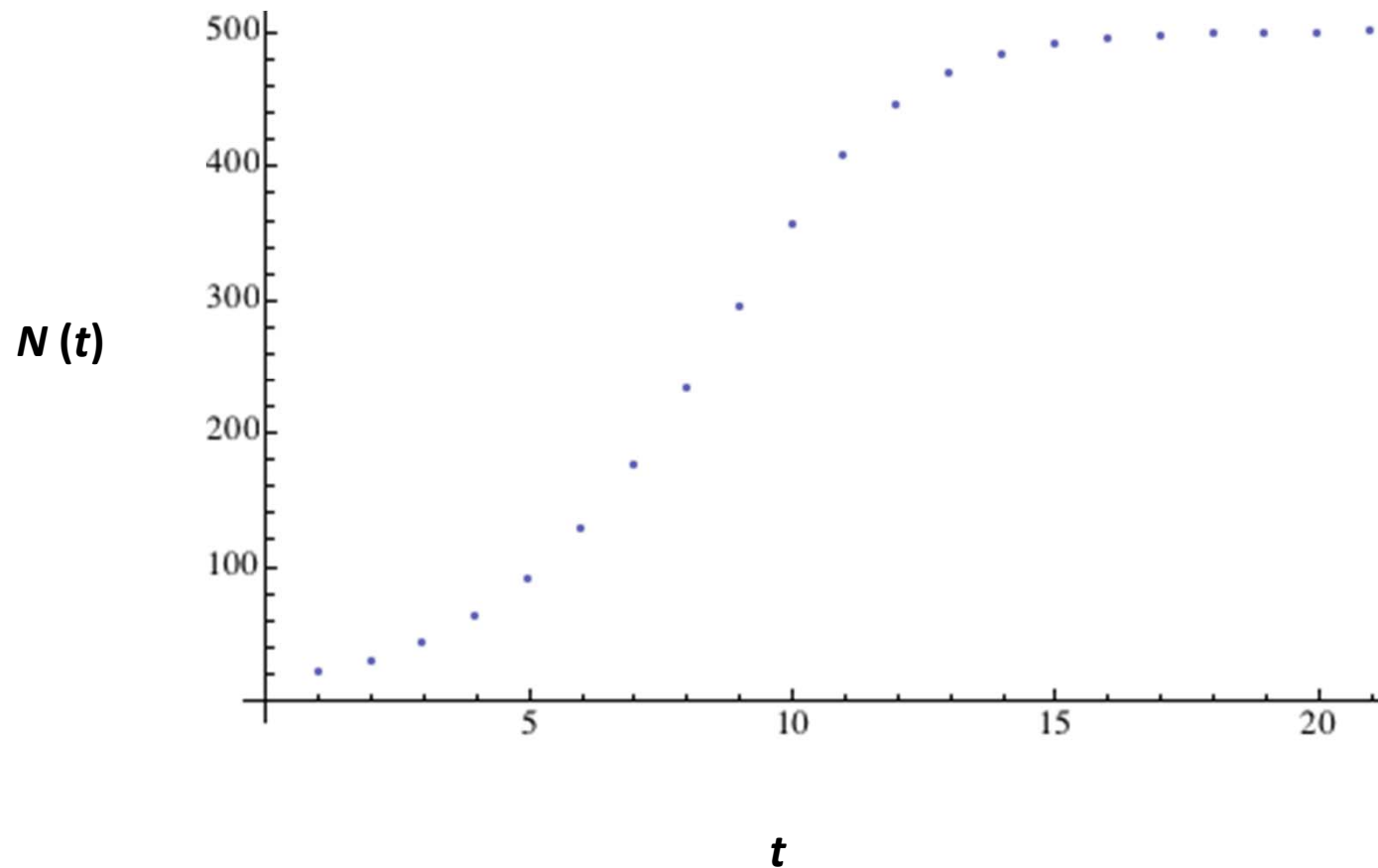
Continuous time

- Processes being studied occur continuously (e.g., population growth)
- Models formulated as differential equations (calculus)
- Methods for solving these have been extensively studied

Understanding models

- Sometimes we don't just want to make predictions for one set of circumstances.
- We may want to understand how the model (equation) behaves under a number of situations.
- “How will y change as we vary x ? Under what conditions will y remain unchanged?”

Population growth: 'The Logistic Growth Curve'



Logistic growth can be characterized
by the following function:

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

This is a differential equation that describes the rate at which a variables changes over time.

Logistic growth can be characterized by the following function:

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

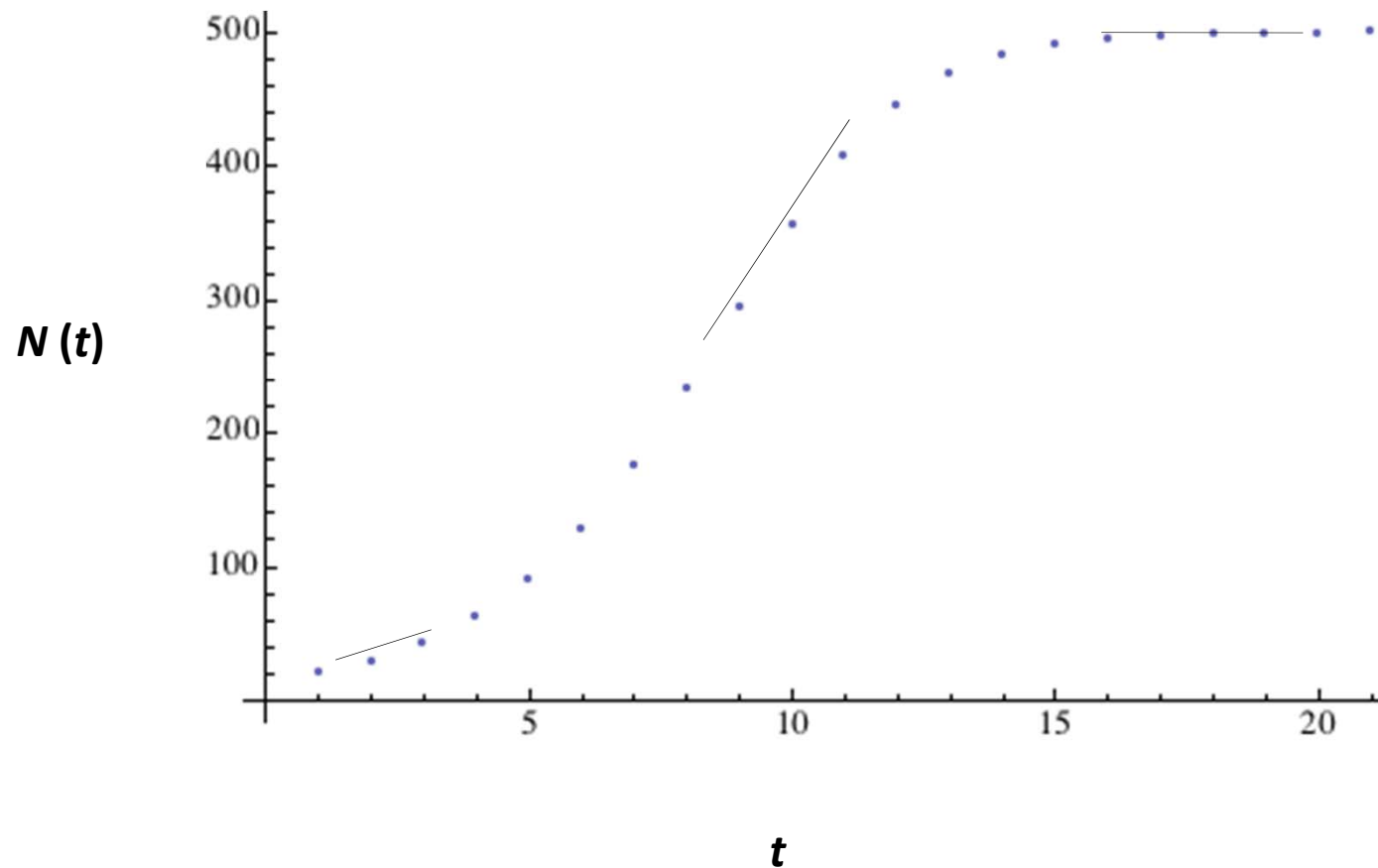
parameter, r ,
growth rate

parameter, K ,
carrying capacity

dN/dt is the rate
of change in
population size
(N) over time (t)

variable, N ,
population size

If we graph the function over time, we can see that **the slope** changes.



Understanding the model

- What does a positive slope mean?
- What does a slope of 0 mean?
- What would we say about a population with a slope of 0?

Finding equilibria

- What are the equilibrium points of this model with respect to N ?
- When does $dN/dt = 0$?

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

What is an equilibrium?

- A system at equilibrium does not change over time (plural *equilibria*). A particular value of a variable is called an equilibrium value if, when the variable is started at this value, the variable *never changes*.
- In a discrete time model, $n(t+1)$ must equal $n(t)$.
- In a continuous time model, dn/dt must equal zero.

Solving an equilibrium

Step 1:

- For a discrete time equation, replace $n(t+1)$ and $n(t)$ with \hat{n} . e.g., the discrete time exponential growth model, $n(t+1) = Rn(t)$ becomes $\hat{n} = R\hat{n}$.
- For a continuous time differential equation, replace $n(t)$ with \hat{n} and set $\frac{dn}{dt} = 0$.
e.g., the continuous time exponential growth model $\frac{dn}{dt} = rn(t)$ becomes $0 = r\hat{n}$.

Step 2: Solve the equilibrium condition for \hat{n} . When canceling a term from both sides of an equilibrium condition, check if that term could equal zero for some value of \hat{n} . If so, that value of \hat{n} is an equilibrium term of the model.

Step 3: Check each equilibrium by plugging it back into the original dynamical equation and confirming that the system remains constant. Also check that each equilibrium is biologically valid (e.g., is non-negative if the variable represents the number of individuals).

Finding equilibria

When N is 0, dN/dt is 0.

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

When r is 0, dN/dt is 0.

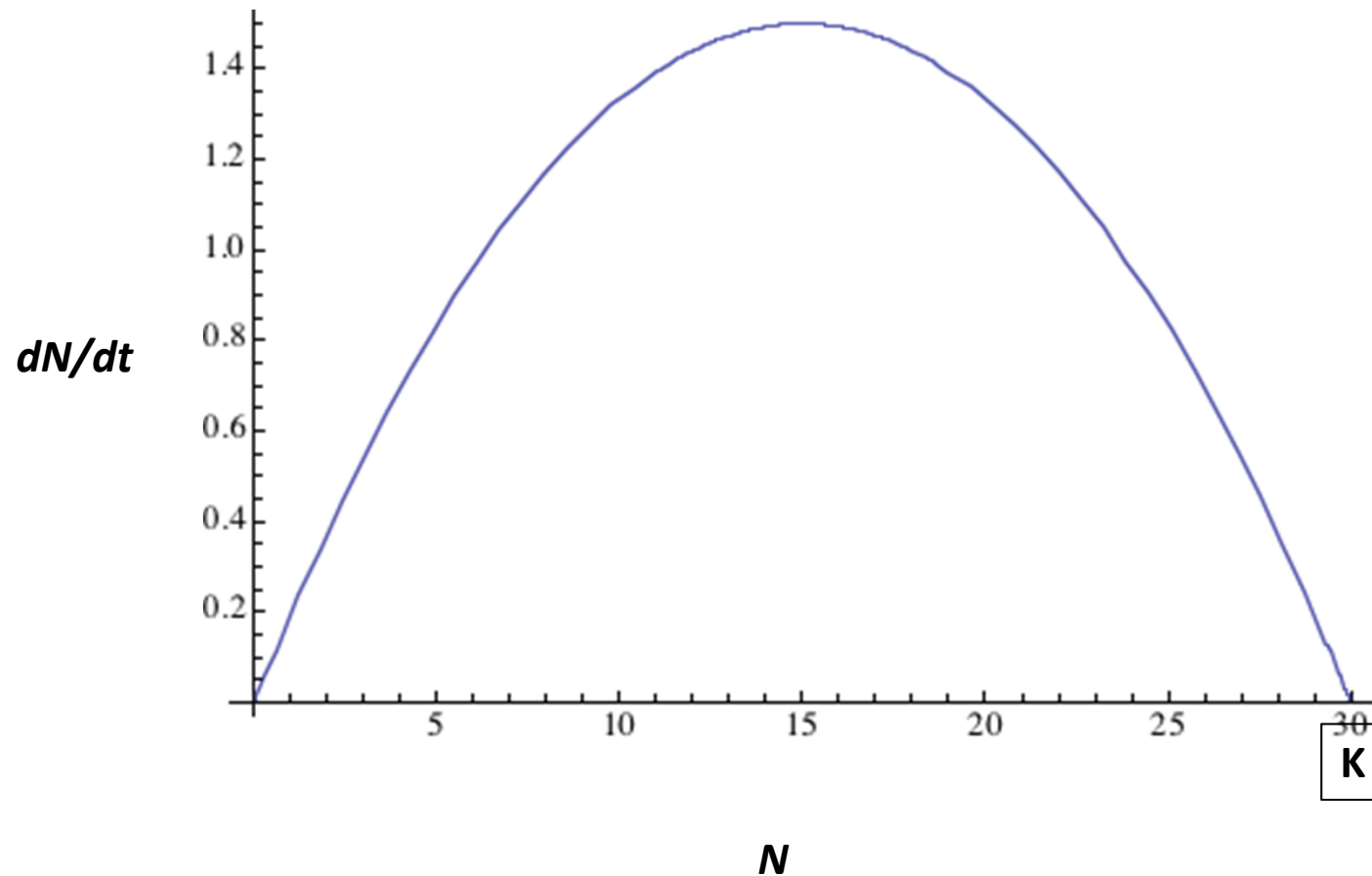
$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

When $N = K$, dN/dt is 0.

$$\frac{dN}{dt} = rN \left(\frac{K - N}{K} \right)$$

Graphical analysis

- What does it mean if the function crosses 0?



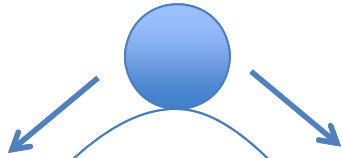
Graphical analysis

- Graphing an equation can tell us a lot about the model. We can learn:
 - Where are the equilibria (no change in X , $\Delta X = 0$).
 - How stable the equilibria are (more on that in a few slides).
 - How variable X changes through time or as other conditions vary.

Stability

- An equilibrium is *locally stable* if a system near the equilibrium approaches it (i.e. it is *attracting*).
- An equilibrium is *unstable* if a system near the equilibrium moves away from it (i.e. it is repelling)

unstable



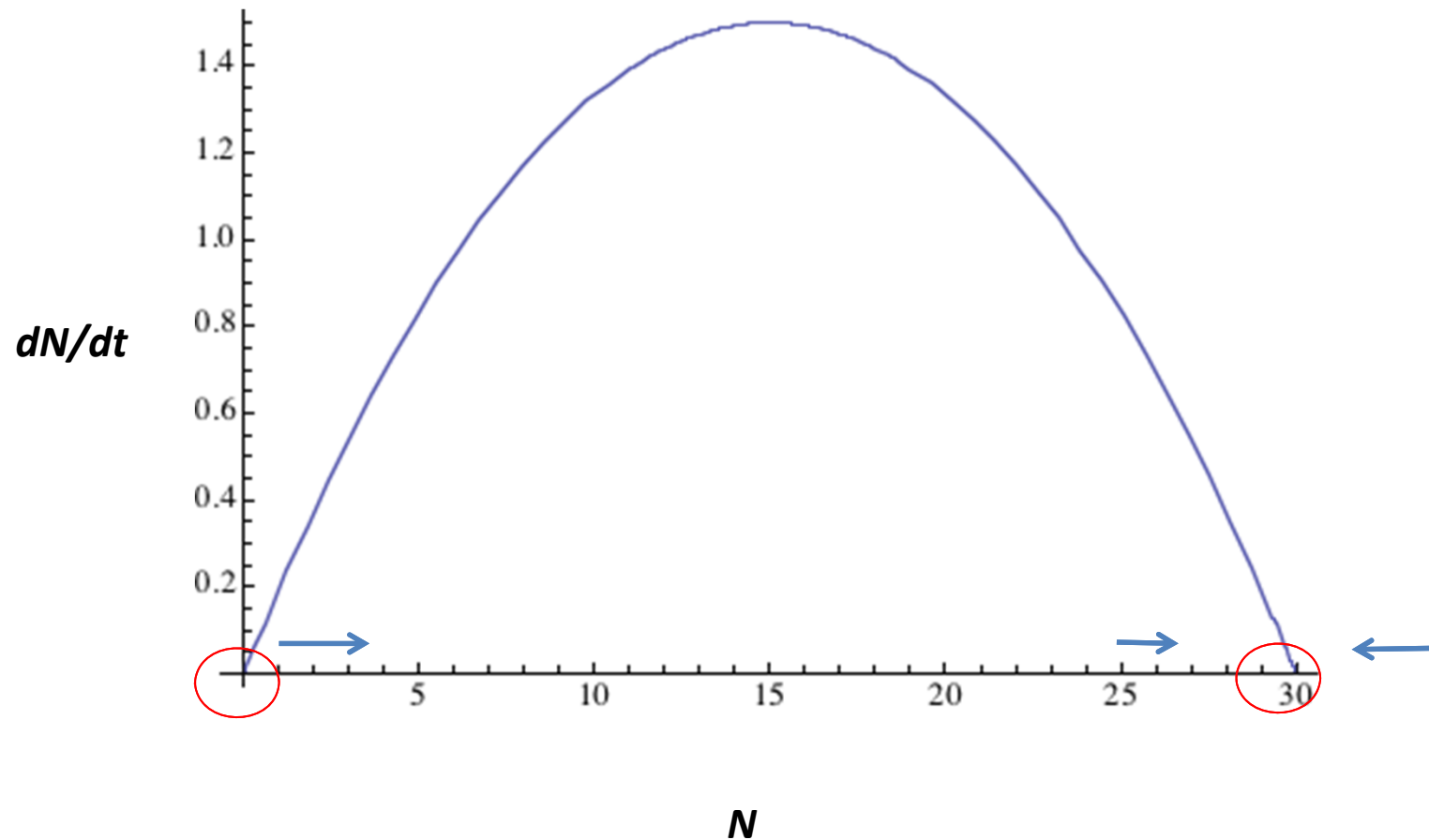
Locally
stable



Stability

- Does the system move toward the equilibrium?
 - If yes, it is a stable equilibrium.
 - If no, it is an unstable equilibrium.

What happens as we perturb the system away from these two equilibria?



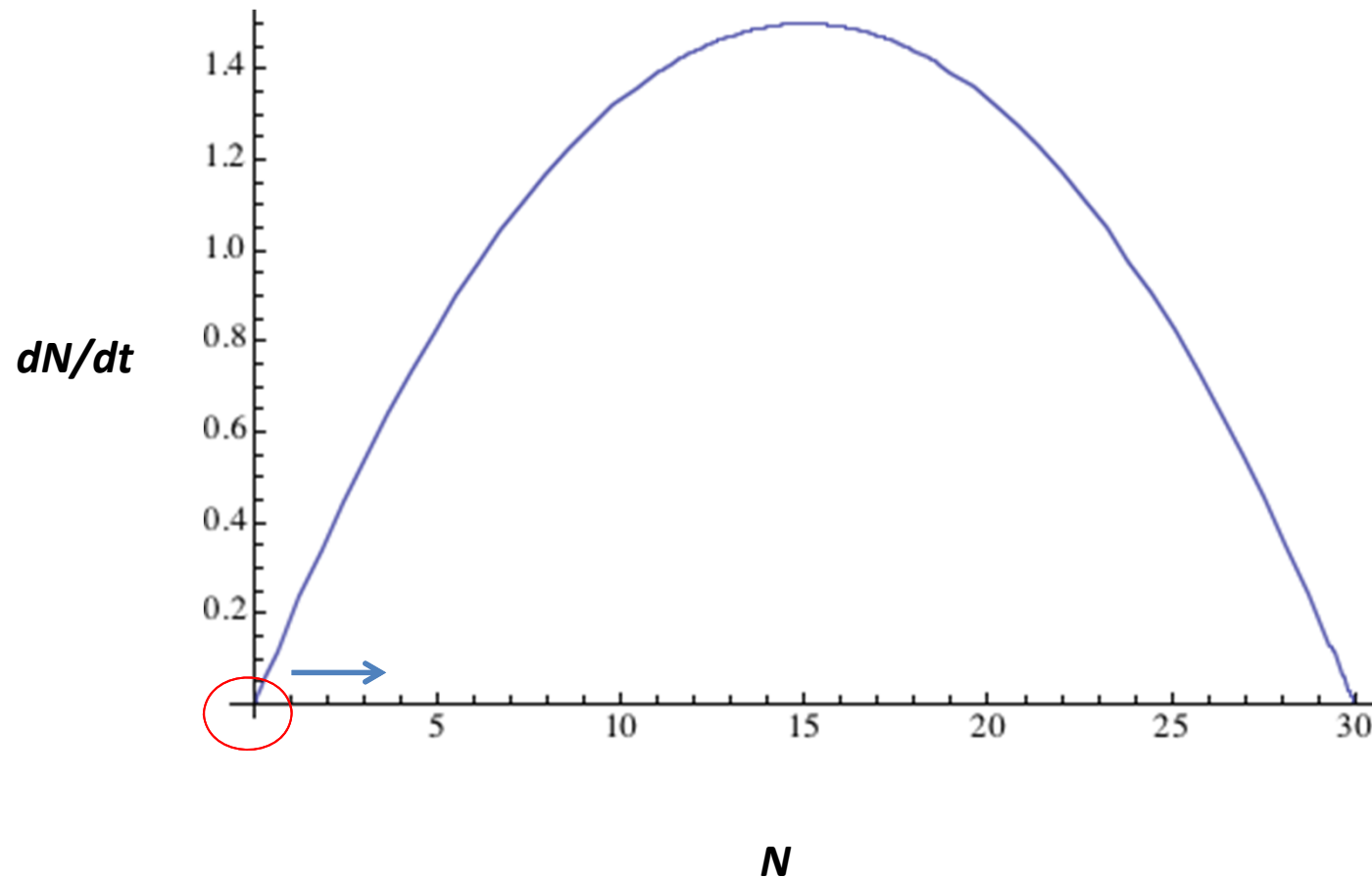
Behaviour near $N=0$

$$N = 0 \quad \frac{dN}{dt} = r(0) \left(\frac{30 - 0}{30} \right) = 0$$

$$N = 1 \quad \frac{dN}{dt} = r(1) \left(\frac{30 - 1}{30} \right) = 0.97r$$

$$N = 2 \quad \frac{dN}{dt} = r(2) \left(\frac{30 - 2}{30} \right) = 1.87r$$

What happens as we perturb the system away from these two equilibria?



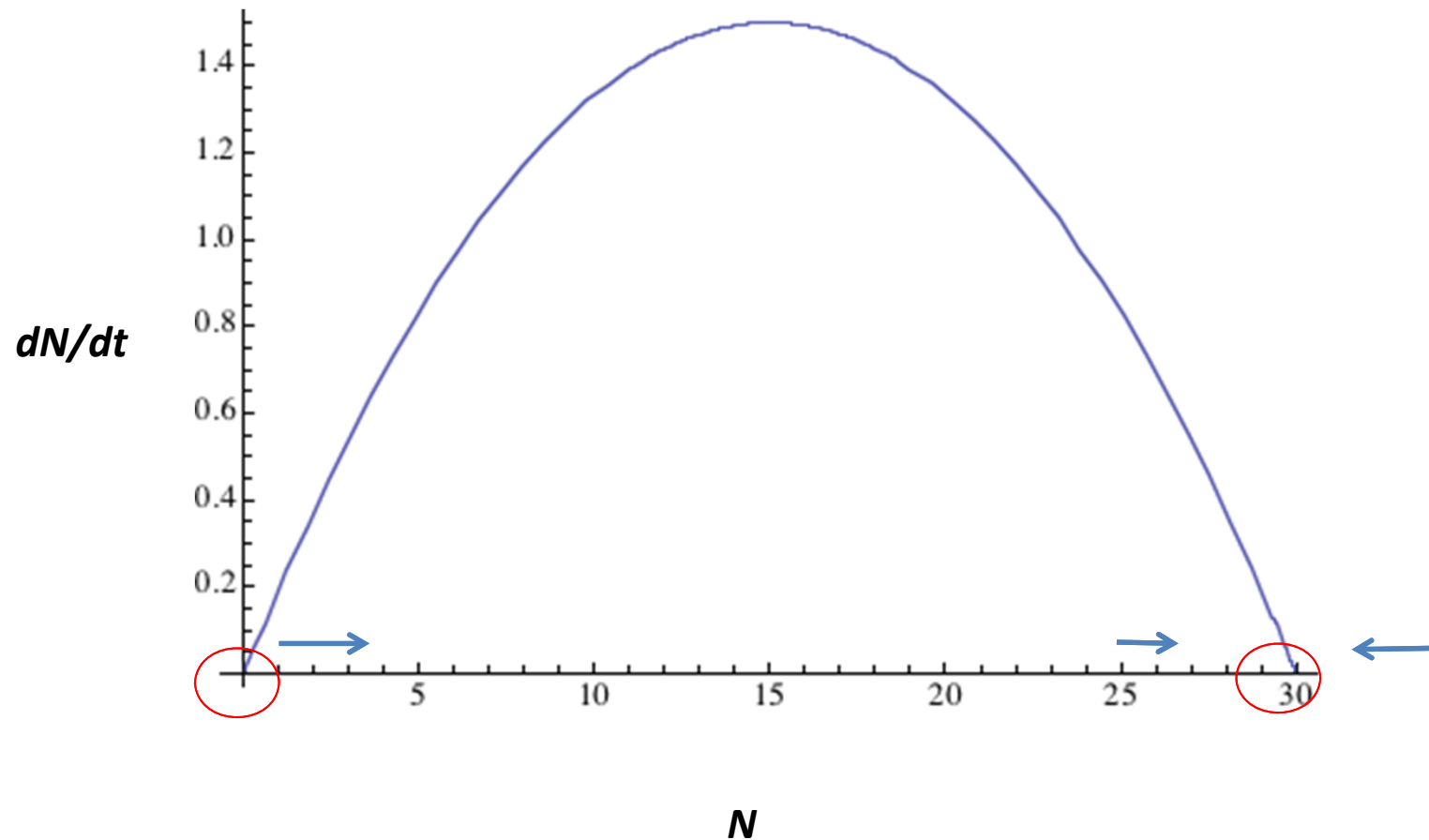
Behaviour near $N = K$

$$N = 30 \quad \frac{dN}{dt} = r(30) \left(\frac{30 - 30}{30} \right) = 0$$

$$N = 29 \quad \frac{dN}{dt} = r(29) \left(\frac{30 - 29}{30} \right) = 0.97r$$

$$N = 31 \quad \frac{dN}{dt} = r(31) \left(\frac{30 - 31}{30} \right) = -1.03r$$

What happens as we perturb the system away from these two equilibria?



Mathematical modeling

- We will be using a number of models in this class and parameters such as equilibria and stability have important biological meaning.
- We will go through each model as it comes up, but these are the basic skills you'll need to understand them.
- You won't be expected to make your own models, only understand the ones shown to you.

Simulations

- Sometimes general solutions don't exist (especially when the models become more complex).
- Computer simulations can be used to track changes in allele frequencies (or other variables) over time under scenarios involving different values of particular parameters, and to determine whether an equilibrium is reached
- They can provide insight into the dynamics of complex equations

Short Introduction to Genetic Variation

Learning outcomes

By the end of this section you should:

- Understand why genetic variation is important
- Be comfortable with simple terminology (e.g. loci, gene, allele, mono and polymorphic, identify by state and descent)
- Be familiar with two different ways of quantifying genetic variation (prop. of polymorphic loci and average heterozygosity)
- **Be able to calculate genotype frequencies from absolute numbers of different genotypes, and allele frequencies from genotype frequencies.**

Genetic Variation

Refers to the existence of DNA sequence variation among a set of alleles at a given locus in a population.

Almost all evolutionary processes require genetic variation to have any affect:

- natural selection
- genetic drift
- gene flow/migration
- Inbreeding/assortative mating

Which one doesn't???

Example: natural selection

- Natural selection is the differential survival and/or reproductive success of individuals differing in phenotype, and is the only evolutionary process that produces adaptation
- Imagine a phenotypically homogeneous population.



- Can natural selection occur in this population?

Aside: what is a population?

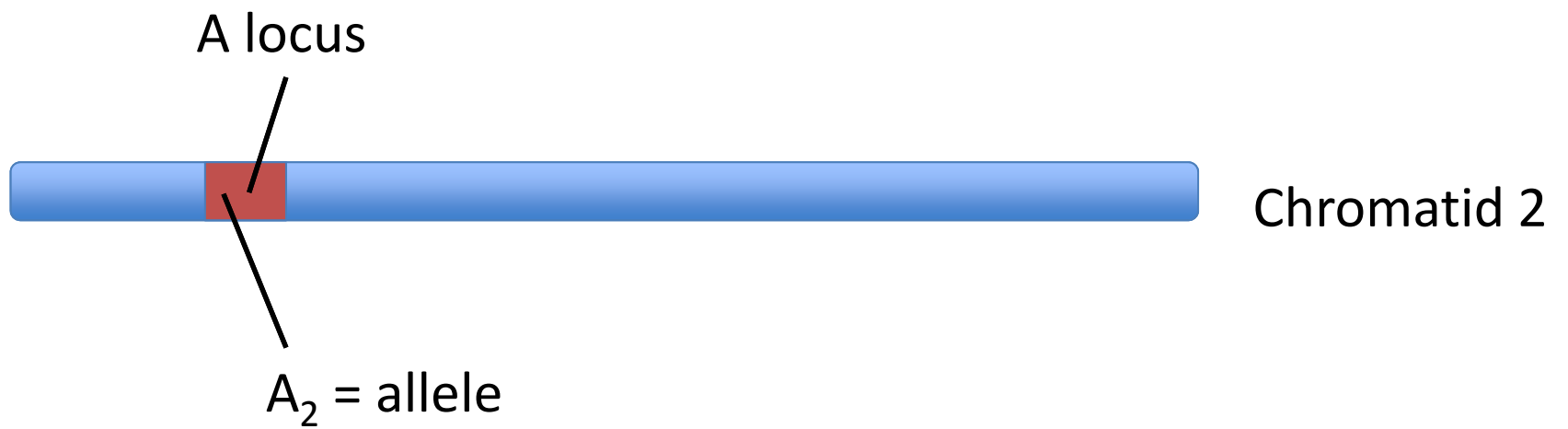
- consists of members of the same species
- Are usually geographically *continuous*
- any individual can potentially mate with any other individual in the population
- can be studied as a unit; for example, you can study the size of a population, or the frequency of a particular allele in that population

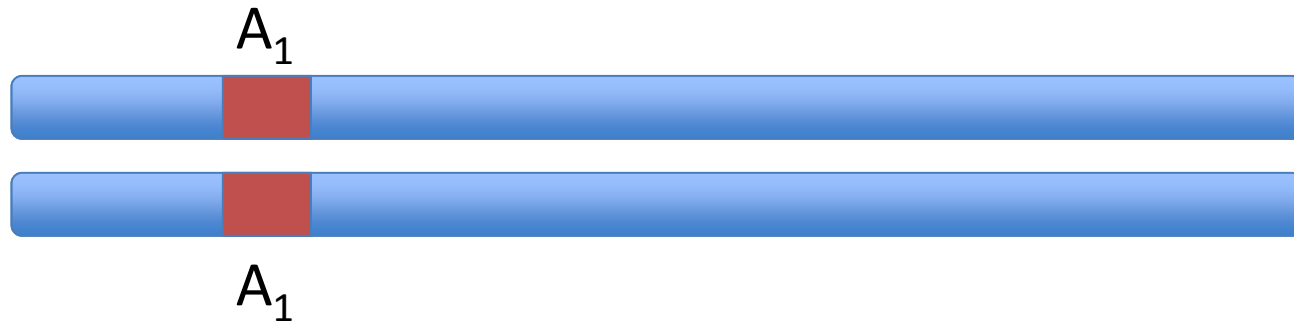
Population genetics

- Seeks to understand the structure and dynamics of naturally occurring genetic variation both within and among populations/species

Some definitions

- Locus – any location (sequence of DNA) on a chromosome
- Gene – a sequence of DNA that codes for a protein
- Allele – the alternative form(s) of DNA at a particular locus
- Monomorphic locus – only one allele for this locus in the population
- Polymorphic/segregating locus – a locus with multiple alleles in the population





Monomorphic
loci



Polymorphic
loci

Genotype and allele frequencies

- Let's call our two alleles A_1 and A_2
- Three possible (combinations) **genotypes** exist, A_1A_1 , A_1A_2 , and A_2A_2 .
- Of course, A_1A_2 and A_2A_1 are the same genotype. Note that there are *two distinct* forms (A_1 sperm and A_2 egg and A_1 egg and A_2 sperm) – it will be obvious later why you should care.

Notation: Genotype Frequencies

Genotype	A_1A_1	A_1A_2	A_2A_2
Genotype frequency (P , not p)	P_{11}	P_{12}	P_{22}
Hamilton's terminology	X	Y	Z

- The sum of all genotype frequencies for a given locus must be one:

$$P_{11} + P_{12} + P_{22} = 1$$

(because they represent mutually exclusive combinations of the alleles as each individual can only have one genotype)

Genotype frequencies can be calculated from counts of the number of individuals of each genotype:

$$P_{11} = \frac{N_{11}}{N}$$

where $N_{11} + N_{12} + N_{22} = N$

- P_{11} is a frequency, but can also be thought of as the probability of choosing a random individual from the population with the genotype A_1A_1 (e.g. $P_{11} = \Pr(X = A_1A_1)$) in a random draw from the population

Allele frequencies

- How many A_1 alleles are in an A_1A_1 individual?
- How many in an A_1A_2 individual?

You must account for the number of a particular allele in different genotypes when calculating allele frequencies from genotype frequencies!

Genotype to Allele Frequencies

- The frequency of the A_1 allele in the population (p) is:

$$p = P_{11} + \frac{1}{2} P_{12}$$

or

$$p = \frac{2N_{11} + N_{12}}{2N} \quad \text{where } N = N_{11} + N_{12} + N_{22}$$

2N alleles in the population (gene pool) as individuals are diploid

- Again, be careful about P vs. p
- What is the frequency of the A_2 allele?

Answer

- The frequency of the A_2 allele in the population is:

$$q = 1 - p = P_{22} + \frac{1}{2} P_{12} = \frac{N_{12} + 2N_{22}}{2N}$$

- Because p and q are mutually exclusive frequencies,

$$p + q = 1$$

Two ways to think about p and q

- p is the frequency of A_1 alleles among all of the alleles at this locus in the population (of which there are two types in this example)
- Or equivalently, p is the probability that an allele picked at random at this locus in the population is A_1 ($p = \Pr(X = A_1)$)

- How would you calculate p_1 from genotype frequencies if there were three alleles at the locus of interest?
- Can you find a general equation for finding p regardless of the number of alleles present (given all genotypic frequencies)?

3 alleles

$$p = (P_{11} \times 1) + \left(P_{12} \times \frac{1}{2}\right) + \left(P_{13} \times \frac{1}{2}\right)$$

General equation for loci with more than two alleles

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}$$

Where P_{ii} is the frequency of the homozygous genotype for allele in question and P_{ij} are the frequencies of the heterozygous genotypes containing this allele.

Frequency of alkaline phosphatase genotypes in a sample of English people (Harris data)

Genotype	Number
SS	141
SF	111
FF	28
SI	32
FI	15
II	5
Total	332

Exercise

- Calculate the frequency of the **three** alkaline phosphatase alleles (S, F, and I) in the English population.

Genotype	Number
SS	141
SF	111
FF	28
SI	32
FI	15
II	5
Total	332

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} P_{ij}$$

Quantifying Genetic Variation

- 1) Proportion of polymorphic loci in a population

$$= \frac{\text{number of polymorphic loci}}{\text{total number of loci examined}}$$

Example – Table 2.1. 6 loci with 10 samples

Locus	AA	Aa	aa	p	Polymorphism
A	10	0	0	1.0	monomorphic
B	3	4	3	0.5	polymorphic
C	6	3	1	0.75	polymorphic
D	1	1	8	0.15	polymorphic
E	0	0	10	0	monomorphic
F	10	0	0	1.0	monomorphic

3/6 loci are polymorphic, therefore
proportion polymorphic = 0.5

Quantifying Genetic Variation

2) Heterozygosity is the frequency of heterozygotes. Average heterozygosity is the frequency of heterozygotes averaged over the n sampled loci:

$$\overline{H}_{obs} = \frac{1}{n} \sum_{i=1}^n H_i$$

This is a more informative means of quantifying genetic diversity in a population.

Example – Table 2.1

Locus	AA	Aa	aa	p	Polymorphism	Heterozygosity
A	10	0	0	1.0	M	0
B	3	4	3	0.5	P	0.4
C	6	3	1	0.75	P	0.3
D	1	1	8	0.15	P	0.10
E	0	0	10	0	M	0
F	10	0	0	1.0	M	0

$$\bar{H} = 0.133$$