Lecture 1: Probability and Review of Chapter 1

Announcements:

1. Office hours MWF for 1h after class or by appointment, 432 Animal Sciences Building

2. Please begin reading textbook, ch. 1 and 2. Let me know if you find errors in the textbook

3. Compass is now set up for this course, with lectures and handouts available. Please let me know if you cannot access it.

Review of topics

- **Gene**: unit of inheritance transmitted from parents to offspring

- **Locus** (plural loci): chromosomal position of a gene (note: most of these terms *also apply to non-coding regions*)

- **Alleles**: the different forms of a gene

- **Diploid**: genes present in two copies, one from each parent

- **Genotype**: the two alleles of a single gene present in an individual

- **Homozygote**: two alleles in an individual are the same
- **Heterozygote**: two alleles in an individual are different
- **Polyploid**: Carrying more than two copies of each gene (e.g., wheat is hexaploid)
- In placental mammals, **autosomal** chromosomes (non-sex-determining) are diploid, while the X-chromosome is **haplo-diploid**.
- **Haploid**: Carrying only one copy of each gene (e.g., mitochondrial DNA, Y-chromosome, gametes)
- Y-chromosome is paternally inherited; mtDNA and chloroplast DNA *generally* show maternal inheritance

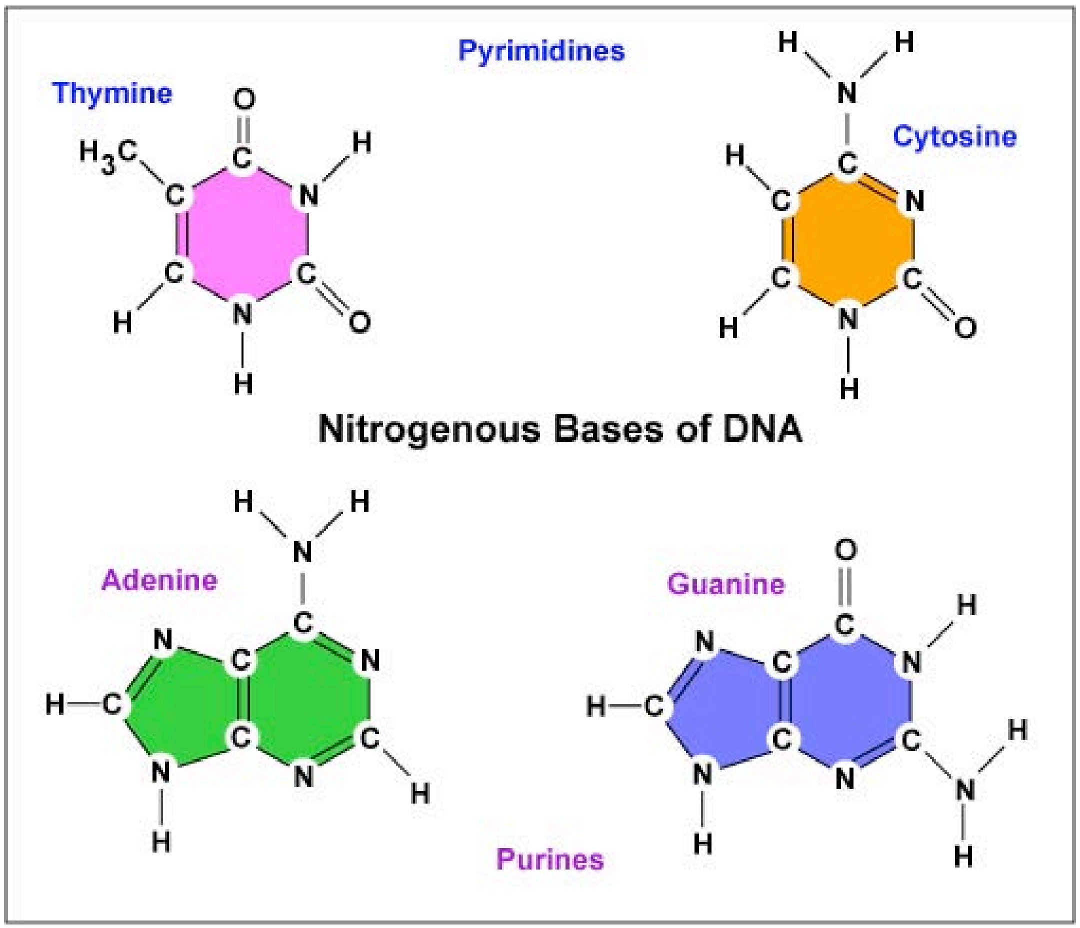
- **Haplotype**: an array of linked genes or alleles on a particular copy of a chromosome. Haplotypes may be disrupted by recombination.

- **Recombination**: The trading of fragments of genetic material between chromosomes before the egg and sperm cells are created, usually the breaking and rejoining of homologous chromosomes.

- Recombination between genes increases with increasing physical separation. Rule of thumb: 1 **map unit** or 0.01 recombination between two genes represents approximately 1 million nucleotides separation.

- **Wild type** alleles: result in a wild or normal phenotype.

- DNA bases or nucleotides consist of the **purines** adenine and guanine; and the **pyrimidines** cytosine and thymine. **Transitions** are mutations from purine to purine, or pyrimidine to pyrimidine. Mutations from purine to pyrimidine or vice versa are called **transversions**.

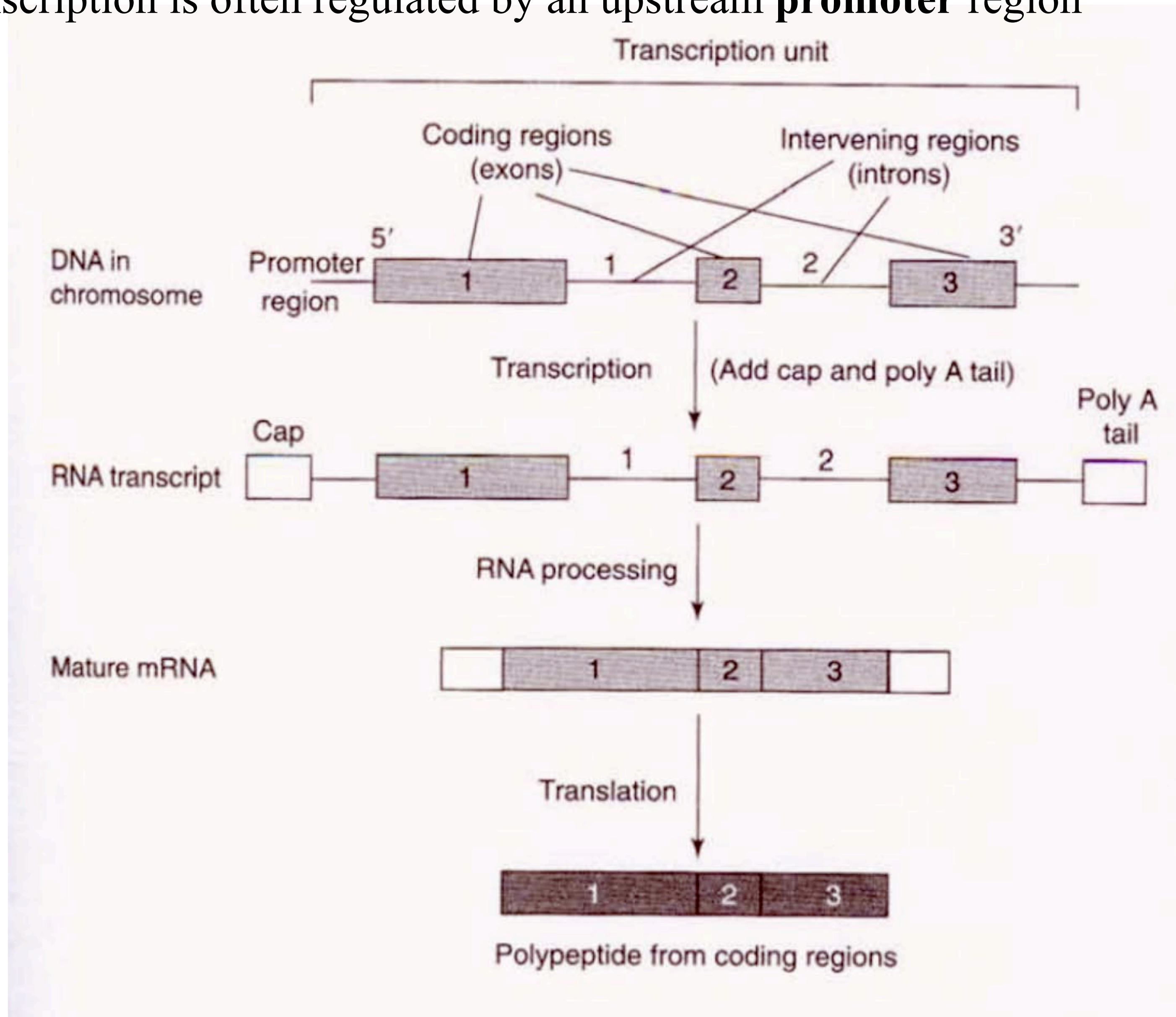


64 **codons** specify 20 amino acids (or "stop"), with considerable two- or four-fold **degeneracy**, especially in the 3d position. Some mutations in a coding region will be silent (**synonymous**); others **non-synonymous**

First position	Second position: U	Second position: C	Second position: A	Second position: G
U	UUU, UUC: Phe (F) UUA, UUG: Leu (L)	UCU, UCC, UCA, UCG: Ser (S)	UAU, UAC: Tyr (Y) UAA, UAG: Stop	UGU, UGC: Cys (C) UGA Stop UGG Trp (W)
C	CUU, CUC, CUA, CUG: Leu (L)	CCU, CCC, CCA, CCG: Pro (P)	CAU, CAC: His (H) CAA, CAG: Gln (Q)	CGU, CGC, CGA, CGG: Arg (R)
A	AUU, AUC, AUA: Ile (I) AUG Met (M)	ACU, ACC, ACA, ACG: Thr (T)	AAU, AAC: Asn (N) AAA, AAG: Lys (K)	AGU, AGC: Ser (S) AGA, AGG: Arg (R)
G	GUU, GUC, GUA, GUG: Val (V)	GCU, GCC, GCA, GCG: Ala (A)	GAU, GAC: Asp (D) GAA, GAG: Glu (E)	GGU, GGC, GGA, GGG: Gly (G)

Genes consist of coding regions called **exons**, separated by non-coding **introns**. They may also have 5'- and/or 3' untranslated regions (**UTRs**).

Transcription is often regulated by an upstream **promoter** region

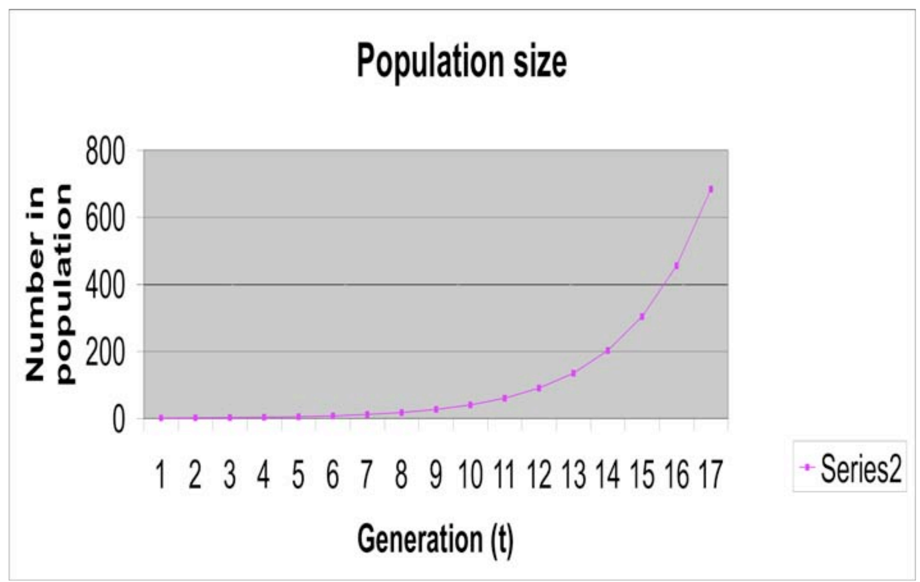


Model to predict the number of individuals in a population (p15)

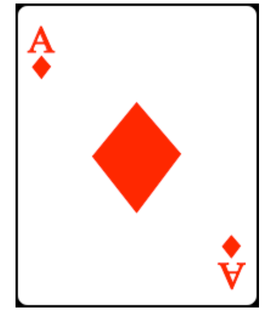
- N_t = number in generation t
- N_{t+1} = number in generation t + 1
- R = **net replacement rate** = N_{t+1}/ N_t
- $N_{t+1} = R\ N_t$
- $N_{t+2} = R\ N_{t+1}$
- $N_{t+2} = R^2\ N_t$
- $N_t = R^t\ N_0$ = population size in generation t

Δ (delta) or difference (p16)

- $\Delta N = N_{t+1} - N_t$
- $\Delta N = R\,N_t - N_t$
- $\Delta N = N_t\,(R - 1)$
- If R = 1, then the population size does not change.
- Example R=1.5



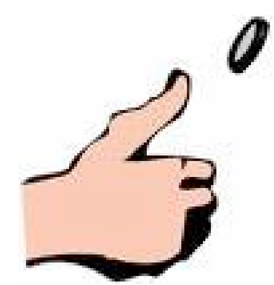
Probability Intro (p22)



- P(A) = probability of event A occurring
- $0 \leq P(A) \leq 1$
- P(A) = m/n = relative frequency
  - where m = number of outcomes of event A
  - and n = number of possible outcomes.
- If A can not occur, m=0 and P(A) = 0.
- If A must occur, m=n and P(A) = 1.

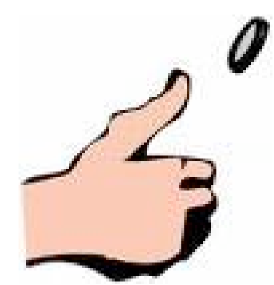
Mutually exclusive

- Events A and B are mutually exclusive if the occurrence of one precludes the occurrence of the other.
- Either event A or event B may occur but not both at the same time.
- For example, a coin tossed a single time can land with the head up, or the tail up, but not both up.



Complement

- Complement of A is 'not A'.
- P(A) + P(not A) = 1
- P(not A) = 1- P(A)
  - Because A and 'not A' are mutually exclusive events.
- **Example**: "heads" is the complement of "tails" in a coin toss



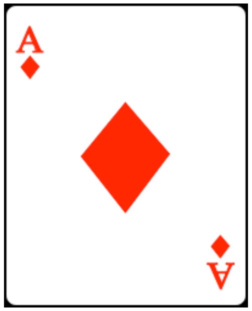
Example 1

- Consider one locus with two alleles (A_1 and A_2) where A_1 is dominant to A_2.
- What is the probability of an offspring with dominant phenotype (A_1-) from a mating of A_1A_2 x A_1A_2?
- Four genotypes are equally likely from this mating: A_1A_1, A_1A_2, A_2A_1 and A_2A_2.

Example 1 continued

- $P(A_1\text{-}) = P(A_1A_1) + P(A_1A_2) + P(A_2A_1)$
- $P(A_1\text{-}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$
- Recessive phenotype and the dominant phenotype are “mutually exclusive”
- $P(A_2A_2) = 1 - P(A_1\text{-}) = 1 - \frac{3}{4} = \frac{1}{4}$
- $P(A_1\text{-}) + P(A_2A_2) = \frac{3}{4} + \frac{1}{4} = 1$

Independent

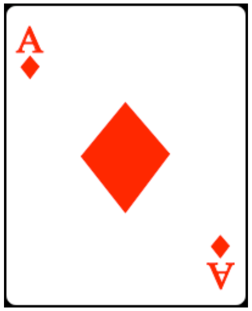


- If the probability of event A is the same whether or not event B occurs, the events are independent.
- Example:

  A: probability of getting an ace (1in 13)

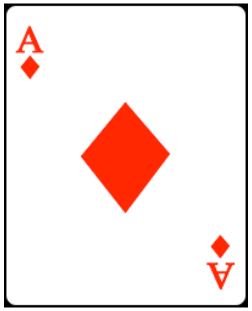
  B: probability of getting diamonds (1 in 4)

Independent



- If the probability of event A is the same whether or not event B occurs, the events are independent.
- If $P(A|B) = P(A)$, then $P(AB) = P(A)P(B)$.
- Also, if events A and B are independent, then $P(AB) = P(A)P(B)$.

Additive law



- P(A or B) = P(A) + P(B) – P(AB)
- Probability of event A occurring or event B occurring is the sum of the probability of event A occurring plus the probability of event B occurring minus the probability that both event A and event B occurred.
- Example: getting an Ace OR Diamonds

Additive rule for more than two events

- P(A or B or C) = P(A) + P(B) + P(C) – P(AB) – P(AC) – P(BC) + P(ABC).

- P(A or B or C or D) = P(A) + P(B) + P(C) + P(D) – P(AB) – P(AC) – P(AD) – P(BC) –P(BD) – P(CD) + P(ABC) + P(ABD) + P(ACD) + P(BCD) – P(ABCD).

Additive rule for mutually exclusive events

- P(A or B) = P(A) + P(B)

- P(A or B or C) = P(A) + P(B) + P(C)

Example: rolling one die
and getting “3” *or* “4” *or* ‘5”

- P(A or B or C or D)
  = P(A) + P(B) + P(C) + P(D)

Conditional probability

- the probability of some event A, given the occurrence of some other event B

- $P(A|B)$ is "the probability of A given B"

- $P(A|B) = P(A \cap B) / P(B)$ for $P(B) \neq 0$.

Conditional probability: Multiplicative law

- $P(A \text{ and } B) = P(A) \times P(B|A)$

Binomial Probability

- $Pr(i) = ((N!/(i!j!))\ p^i\ q^j$
  where $N = i + j$ = total number of events,
  $p = P(A)$ = probability of event A,
  $q = P(B)$ = probability of event B,
  i = number of occurrences of event A
  j = number of occurrences of event B

n! represents the factorial of n, or 1 X 2 X 3… X n
for example, 6! = 1 X 2 X 3 X 4 X 5 X 6 =720
Note: 0! = 1

Multinomial probability

- $Pr(i,j) = (N!/(i!j!k!))\ P^i H^j Q^k$

  where $N = i + j + k$ = total number of events,

  $P = P(A)$ = probability of event A,

  $H = P(B)$ = probability of event B,

  $Q = P(C)$ = probability of event C,

with $P+H+Q = 1$

  i = number of occurrences of event A

  j = number of occurrences of event B

  k = number of occurrences of event C

A few final definitions

- Heuristic: a method used to solve a problem, often "rule of thumb"
- Stochastic: chance events
- Monte Carlo methods: computational algorithms that rely on repeated computation and random or pseudo-random numbers (so not deterministic)