# Today’s Lecture

- Probability models for sequences
- Neutralist vs selectionist interpretations
- Site models
- Comparing models: Likelihood ratios & weight matrices
  - (Hypothesis testing & Neyman-Pearson lemma)

- A *probability space* ($S$,$P$) is a sample space $S$ with a prob dist'n $P$ on $S$.
- Prob dist'n on $S$ is sometimes called a *probability model* for $S$, particularly if several dist'ns are being considered.
  - Write models as $M_1$, $M_2$ , probabilities as $P(s \mid M_1)$, $P(s \mid M_2)$.
  - e.g.
    - $M_1$ = prob dist'n for splice site seqs,
    - $M_2$ = prob dist'n for "background" (arbitrary genomic) seqs.

# Basic Probability Theory Concepts (cont'd)

- An *event* $E$ is a criterion that is true or false for each $s \in S$.
  - defines a subset of $S$ (sometimes also denoted $E$).
  - $P(E)$ is defined to be $\Sigma_{s|E \text{ is true}} P(s)$.
- Events $E_1, E_2, \ldots, E_n$ are *mutually exclusive* if no two of them are true for the same point;
  - then $P(E_1 \text{ or } E_2 \text{ or } \ldots \text{ or } E_n) = \Sigma_{1 \le i \le n} P(E_i)$.
- If $E_1, E_2, \ldots, E_n$ are also *exhaustive*, i.e. every $s$ in $S$ satisfies $E_i$ for some $i$, then $\Sigma_{1 \le i \le n} P(E_i) = 1$.

- For events *E* and *H*, the *conditional probability* of *E* given *H*, is

$$P(E \mid H) \equiv P(E \text{ and } H) \,/\, P(H)$$

(= prob that both *E* and *H* are true, given *H* is true)

  – undefined if $P(H) = 0$.

- *E* and *H* are (*statistically*) *independent* if

$$P(E) = P(E \mid H)$$

(i.e. prob. *E* is true doesn’t depend on whether *H* is true);

or equivalently

$$P(E \text{ and } H) = P(E)P(H).$$

# Probabilities on Sequences

- Let $S$ = space of DNA or protein sequences of length $n$. Possible assumptions for assigning probabilities to $S$:
  - *Equal frequency assumption*: All residues are equally probable at any position;
    - $P(E_r^{(i)}) = P(E_q^{(i)})$ for any two residues $r$ and $q$,
      - where $E_r^{(i)}$ means residue $r$ occurs at position $i$, then
    - Since for fixed $i$ the $E_r^{(i)}$ are mutually exclusive and exhaustive,

$$P(E_r^{(i)}) = 1 / |A|$$

where $A$ = residue alphabet

$$P(E_r^{(i)}) = 1/20 \text{ for proteins, } 1/4 \text{ for DNA).}$$

  - *Independence assumption*: whether or not a residue occurs at a given position is independent of residues at other positions.

- Given above assumptions, the probability of the sequence

  $s = ACGCG$

  (in the space $S$ of all length 5 sequences) is calculated by considering 5 events:

  - Event 1 is that first nuc is A. Probability = .25.
  - Event 2 is that 2d nuc is C. Probability = .25.
  - Event 3 is that 3d nuc is G. Probability = .25.
  - Event 4 is that 4th nuc is C. Probability = .25.
  - Event 5 is that 5th nuc is G. Probability = .25.

  By independence assumption, prob of all 5 events occurring is the product $(.25)^5 = 1/1024$.

  Since $s$ is the only sequence satisfying all 5 conditions, $P(s) = 1/1024$.

- More generally, under equal freq and indep assumptions,

  prob of nuc sequence of length $n$ = $.25^n$,

  prob of protein sequence of length $n$ = $.05^n$

  in the space $S$ of length $n$ sequences.

# Failure of Equal Frequency Assumption for (Real) DNA

- For most organisms, the nucleotide composition is significantly different from .25 for each nucleotide, e.g.:
  - *H. influenza* .31 A, .19 C, .19 G, .31 T
  - *P. aeruginosa* .17 A, .33 C, .33 G, .17 T
  - *M. janaschii* .34 A, .16 C, .16 G, .34 T
  - *S. cerevisiae* .31 A, .19 C, .19 G, .31 T
  - *C. elegans* .32 A, .18 C, .18 G, .32 T
  - *H. sapiens* .29 A, .21 C, .21 G, .29 T

- Note approximate symmetry: A $\cong$ T, C $\cong$ G,
  - even though we’re counting nucs on just one strand.
  - Expect *exact* equality when counting both strands
- Explanation:
  - Although individual biological features may have non-symmetric composition (local *asymmetry*),
  - usually features are distributed approx *randomly* w.r.t. strand,
  - so local asymmetries *cancel*, yielding overall symmetry.

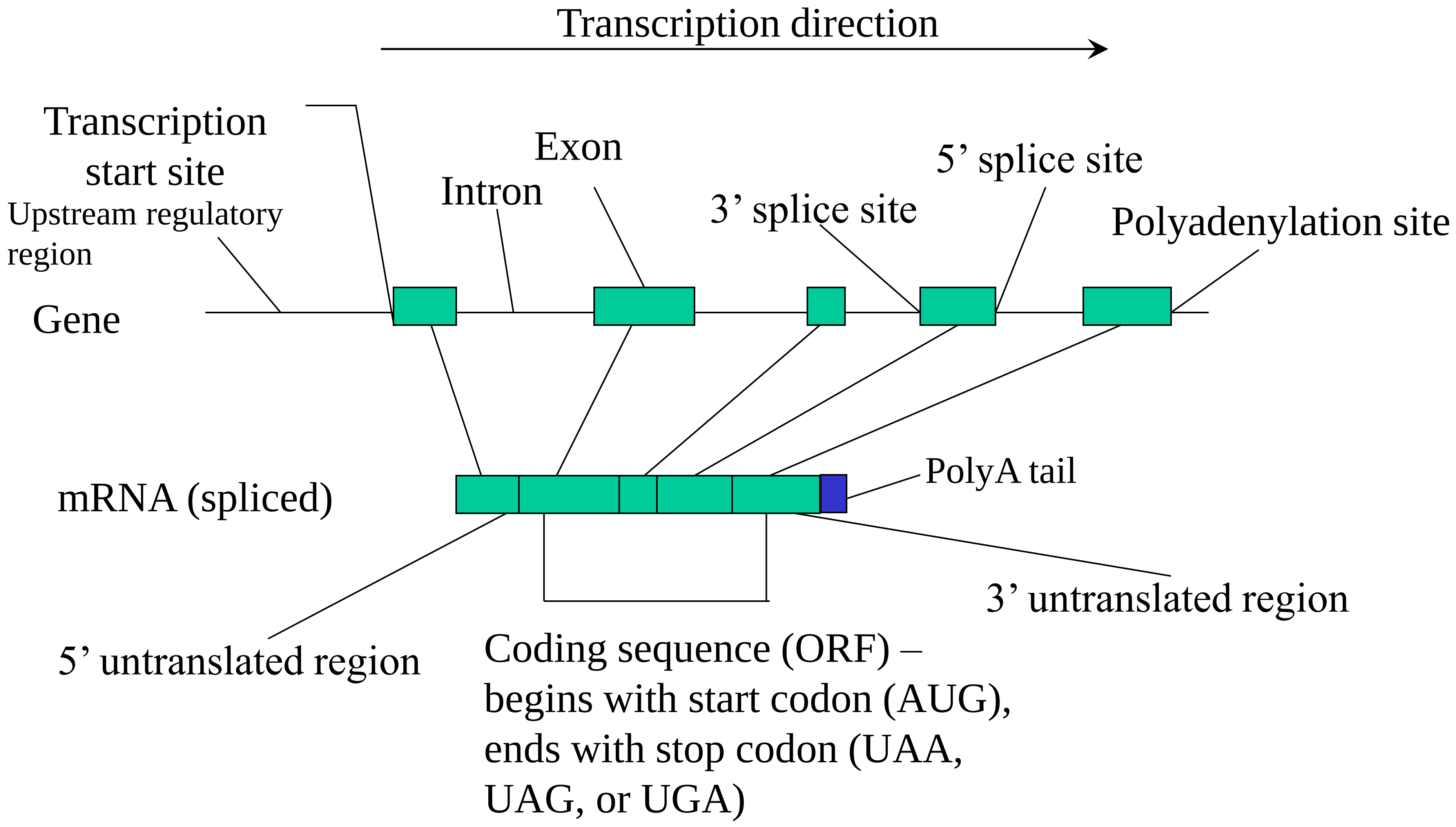
# General Hypotheses Regarding Unequal Frequency

- Neutralist hypothesis:  *mutation bias*
  - e.g. due to nucleotide pool composition
- Selectionist hypothesis: *selection*
  - selection on (many) particular nucleotides
  - selection on mutational bias mechanisms
  - …

# Site Models

- Probability models for short sequences, such as:
  - splice sites
  - translation start sites
  - promoter elements
  - protein “motifs”

# (Protein-coding) Gene Structure in Eukaryotes



Coding sequence (ORF) – begins with start codon (AUG), ends with stop codon (UAA, UAG, or UGA)

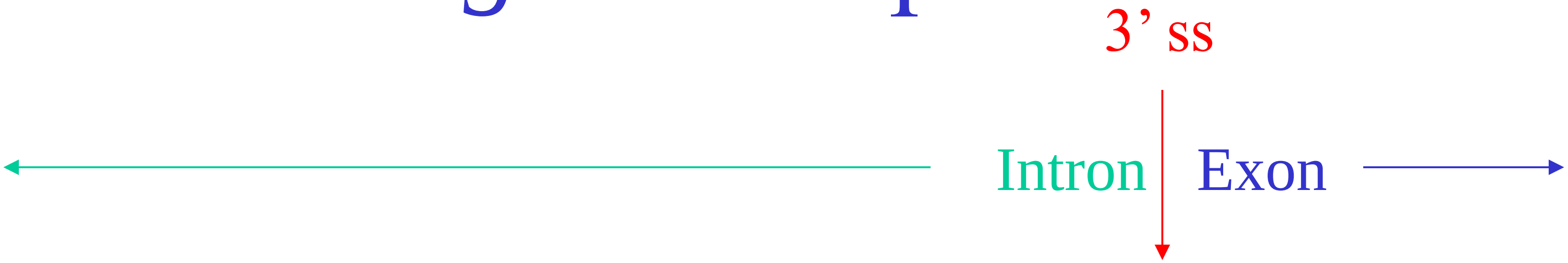
- Assumptions:
  - different examples of site can be aligned *without gaps* (indels) such that tend to have same residues in same positions
  - drop equal freq assumption: allow *position-specific freqs*
  - retain *independence* assumption (for now)

- Applies to short segments (< 30 residues) where
  - precise residue spacing is structurally or functionally important, and
  - certain positions are highly conserved
- Examples:
  - DNA/RNA sequences binding a single protein or RNA molecule
  - Protein internal regions structurally constrained due to folding requirements; or
  - protein surface regions constrained because bind certain ligands

# Construction of Site Models

- Collect examples of site
- Align (without gaps)
- Count occurrences of residues at each position
- Convert to frequencies

# Nucleotide Counts for 8192 *C. elegans* 3’ Splice Sites

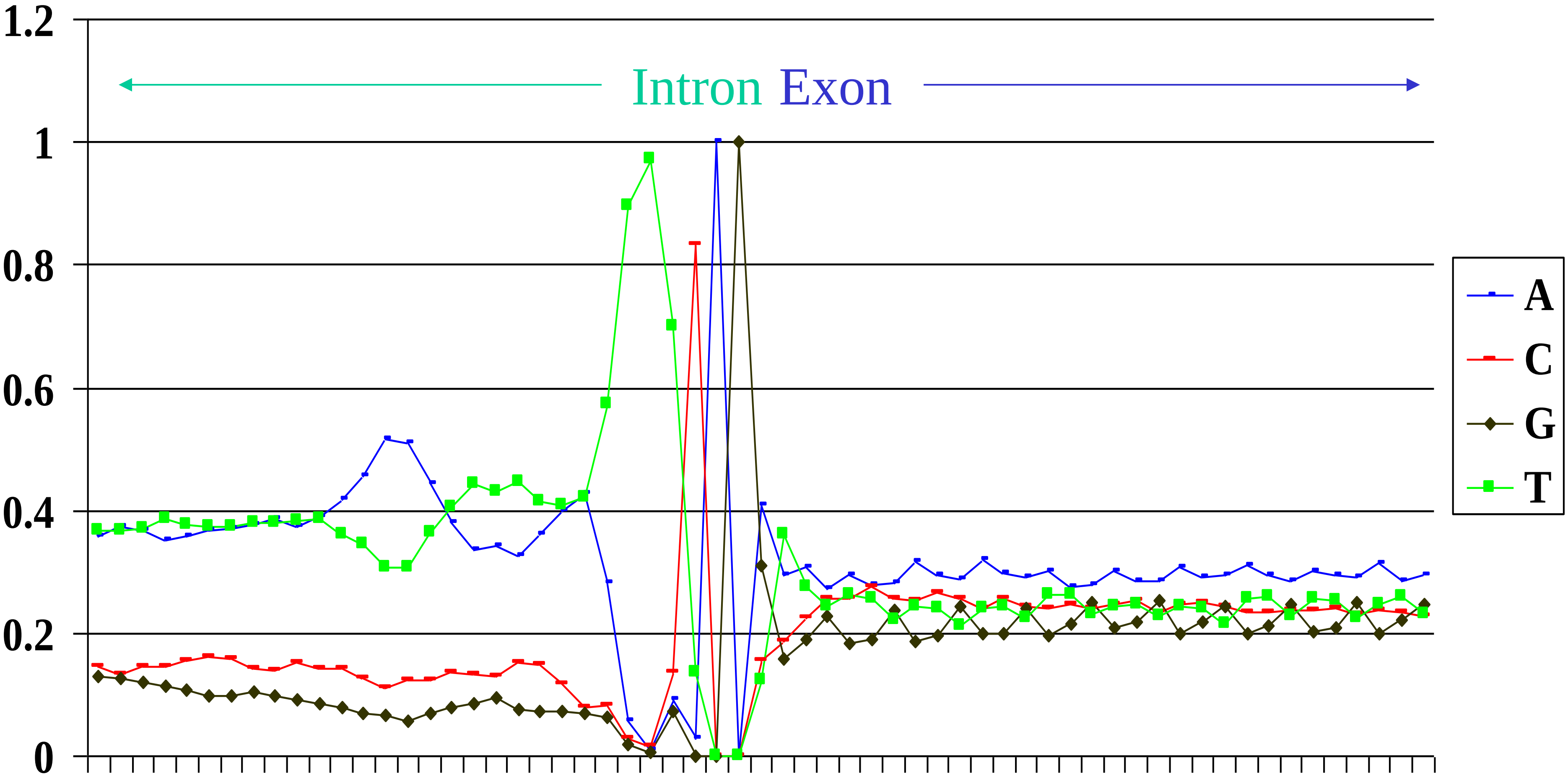

3’ ss

Intron

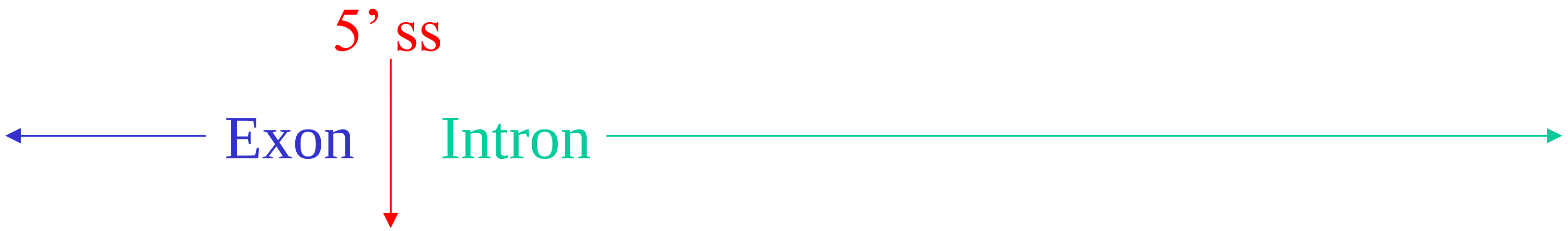
Exon

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 3276 | 3516 | 2313 | 476 | 67 | 757 | 240 | 8192 | 0 | 3359 | 2401 | 2514 |
| C | 970 | 648 | 664 | 236 | 129 | 1109 | 6830 | 0 | 0 | 1277 | 1533 | 1847 |
| G | 593 | 575 | 516 | 144 | 39 | 595 | 12 | 0 | 8192 | 2539 | 1301 | 1567 |
| T | 3353 | 3453 | 4699 | 7336 | 7957 | 5731 | 1110 | 0 | 0 | 1017 | 2957 | 2264 |
| **CONSENSUS** | W | W | W | T | T | t | C | A | G | r | w | w |
| A | 0.400 | 0.429 | 0.282 | 0.058 | 0.008 | 0.092 | 0.029 | 1.000 | 0.000 | 0.410 | 0.293 | 0.307 |
| C | 0.118 | 0.079 | 0.081 | 0.029 | 0.016 | 0.135 | 0.834 | 0.000 | 0.000 | 0.156 | 0.187 | 0.225 |
| G | 0.072 | 0.070 | 0.063 | 0.018 | 0.005 | 0.073 | 0.001 | 0.000 | 1.000 | 0.310 | 0.159 | 0.191 |
| T | 0.409 | 0.422 | 0.574 | 0.896 | 0.971 | 0.700 | 0.135 | 0.000 | 0.000 | 0.124 | 0.361 | 0.276 |

# 3’ Splice Sites – *C. elegans*

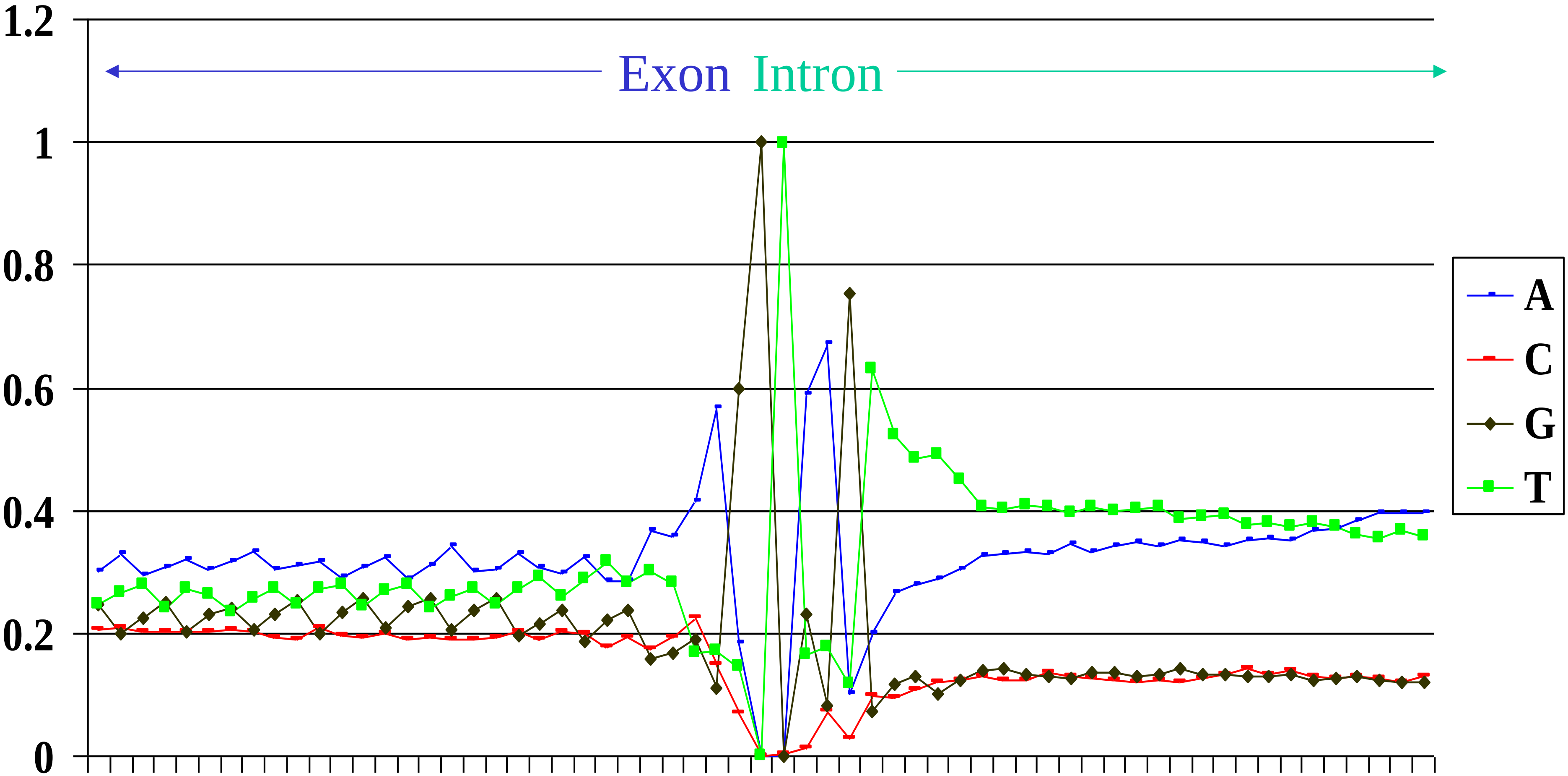


# Nucleotide Counts for 8192 *C. elegans* 5’ Splice Sites



| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 3404 | 4644 | 1518 | 0 | 0 | 4836 | 5486 | 837 | 1632 | 2189 | 2278 | 2355 |
| C | 1850 | 1224 | 583 | 0 | 14 | 118 | 588 | 237 | 801 | 771 | 889 | 986 |
| G | 1562 | 912 | 4891 | 8192 | 0 | 1890 | 672 | 6164 | 589 | 962 | 1056 | 827 |
| T | 1376 | 1412 | 1200 | 0 | 8178 | 1348 | 1446 | 954 | 5170 | 4270 | 3969 | 4024 |
| **CONSENSUS** | x | a | g | G | T | a | a | g | t | t | w | t |
| A | 0.416 | 0.567 | 0.185 | 0.000 | 0.000 | 0.590 | 0.670 | 0.102 | 0.199 | 0.267 | 0.278 | 0.287 |
| C | 0.226 | 0.149 | 0.071 | 0.000 | 0.002 | 0.014 | 0.072 | 0.029 | 0.098 | 0.094 | 0.109 | 0.120 |
| G | 0.191 | 0.111 | 0.597 | 1.000 | 0.000 | 0.231 | 0.082 | 0.752 | 0.072 | 0.117 | 0.129 | 0.101 |
| T | 0.168 | 0.172 | 0.146 | 0.000 | 0.998 | 0.165 | 0.177 | 0.116 | 0.631 | 0.521 | 0.484 | 0.491 |

# 5’ Splice Sites – *C. elegans*



# Conserved Domain in RecR and Class I Topisomerases

```
   RecR   RLAEEKITEVILATNPTVEGEATANYIAELC
   RecM   RLQDDQVTEVILATNPNIEGEATAMYISRLL
   RecR   RVDDVGITEVIIATDPNTEGEATATYLVRMV
   TrsI   IFKENKIDEVIIATDPAREGENIAYKILNQL
   TOP1   KQLAEKADHIYLATDLDREGEAIAWRLREVI
   ORF1   AELLKQANTIIVATDSDREGENIAWSIIHKA
   TOP1   KDALKDADELILATDEDREGKVISWHLLQLL
   TOP1   TIFDKRVKTIILATDAAAEGEYIGRNILYRL
   TOP3   KREARNADYLMIWTDCDREGEYIGWEIWQEA
   TOP3   KRFLHEASEIVHAGDPDREGQLLVDEVLDYL
   RGYR   RNLAVEADEVLIGTDPDTEGEKIAWDLYLAL

CONSENSUS xxxxxxxxxU&uatDxxxEGexxxxxUxxxu
```

*Consensus key*:

Uppercase: all residues chemically similar

lowercase: most are

U,u: bulky aliphatic (I,L,V)

&: bulky hydrophobic (I,L,V,M,F,Y,W)

From RL Tatusov, SF Altschul, and EV Koonin, PNAS 91: 12091-12095

# Probability Models for Sites (assuming independence!)

- For each position $i$, $1 \leq i \leq n$, let $P_i$ be a prob dist'n on the alphabet of residues
  - e.g. constructed using counts at that position in a sample of sites.
  - $P_i(r)$ for each residue $r$ is the probability that $r$ occurs at position $i$ in a sequence.
- Prob dist'n $P$ on the space $S$ of sequences of length $n$ is defined by

$$P(s) = \prod_{1 \leq i \leq n} P_i(s_i)$$

where $s = s_1\, s_2 \ldots s_n$

# Zero Probabilities

- If $P_i(r) = 0$ for some $i$ and $r$, then $P(s) = 0$ for some sequences.
  - may or may not be desirable
- If due to failure to observe residue because of small sample size,
  - should perform “small-sample correction” to change $P_i(r)$ to a small non-zero value.
  - usually done by adding ‘pseudocounts’ to each value in the counts matrix;
    - e.g. add 1 to each cell (has justification in Bayesian statistics)
  - Particularly an issue with proteins, due to larger alphabet size.
- If reflects real biological constraints
  - then leave as 0.
  - e.g. requirement for G at position +1 (first intronic base) in 5’ss

# Comparing Alternative Probability Models

- We will want to consider more than one model at a time, in following situations:
  - To differentiate between two or more hypotheses about a sequence
  - To generate increasingly refined probability models that are progressively more accurate

- First situation arises in testing biological assertion, e.g. “is this a coding sequence?”
  - Compare two models:

1. model associated with a hypothesis $H_{coding}$,
   - assigns each sequence the prob of observing it under expt of drawing a coding sequence at random from genome
2. model associated with a hypothesis $H_{noncoding}$,
   - assigns each sequence the prob of observing it under expt of drawing a non-coding sequence at random

# Likelihood Ratios

- The *likelihood* of a model $M$ given an observation $s$ is

$$L(M \mid s) = P(s \mid M)$$

This is *not* the *probability* of the model! – (the sum over all models is not 1).

- The *likelihood ratio* ($LR$) of two models $M_a$ and $M_0$ is given by

$$LR(M_a, M_0 \mid s) = \frac{L(M_a \mid s)}{L(M_0 \mid s)}$$

The numerator and denominator may both be very small!

- The *log likelihood ratio* ($LLR$) is the logarithm of the likelihood ratio.

# Weight Matrices for Site Models

- LR for sites: (prob under site model) / (prob under non-site (background) model)

$$\frac{P(s \mid M_{\text{site}})}{P(s \mid M_{\text{background}})} = \frac{\prod_{1 \le i \le n} P_i(s_i \mid M_{\text{site}})}{\prod_{1 \le i \le n} P_i(s_i \mid M_{\text{background}})}$$

- LLR = $\sum_{1 \le i \le n} \log(P_i(s_i \mid M_{\text{site}})) - \log(P_i(s_i \mid M_{\text{background}}))$
  - compute by reading from a *matrix* whose *i*-th column contains values $\log(P_i(r \mid M_{\text{site}})) - \log(P_i(r \mid M_{\text{background}}))$ for each residue *r* (with *r* labelling the rows).
    - We use $\log_2$.

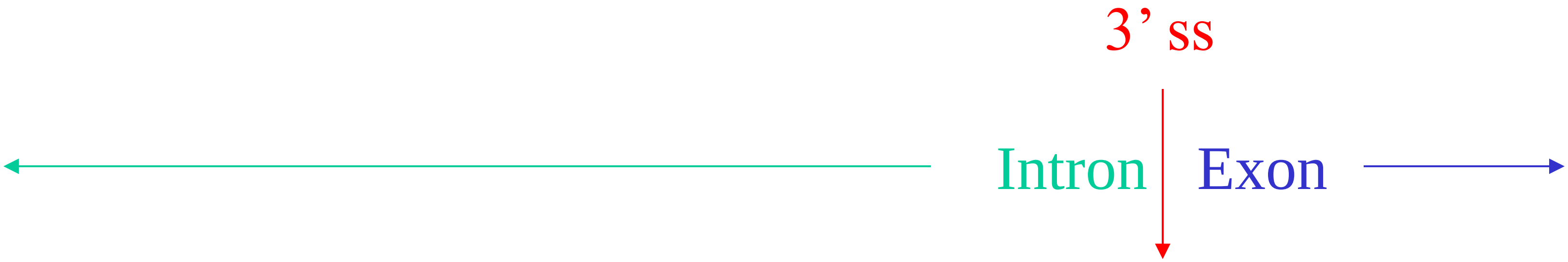
# Example: 3’ splice sites in C. *elegans*

- For *background distribution* take
  - genomic residue freqs computed from C. *elegans* chrom. I:

| | | |
|---|---|---|
| A | 4,575,132: | 0.321 |
| C | 2,559,048: | 0.179 |
| G | 2,555,862: | 0.179 |
| T | 4,582,688: | 0.321 |

  - other choices are possible, e.g. composition of *transcribed regions*
- For the *site distribution* we take
  - site residue freqs from 8192 sites:

# Nucleotide Counts for 8192 *C. elegans* 3’ Splice Sites



3’ ss

Intron

Exon

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 3276 | 3516 | 2313 | 476 | 67 | 757 | 240 | 8192 | 0 | 3359 | 2401 | 2514 |
| C | 970 | 648 | 664 | 236 | 129 | 1109 | 6830 | 0 | 0 | 1277 | 1533 | 1847 |
| G | 593 | 575 | 516 | 144 | 39 | 595 | 12 | 0 | 8192 | 2539 | 1301 | 1567 |
| T | 3353 | 3453 | 4699 | 7336 | 7957 | 5731 | 1110 | 0 | 0 | 1017 | 2957 | 2264 |
| **CONSENSUS** | W | W | W | T | T | t | C | A | G | r | w | w |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.400 | 0.429 | 0.282 | 0.058 | 0.008 | 0.092 | 0.029 | 1.000 | 0.000 | 0.410 | 0.293 | 0.307 |
| C | 0.118 | 0.079 | 0.081 | 0.029 | 0.016 | 0.135 | 0.834 | 0.000 | 0.000 | 0.156 | 0.187 | 0.225 |
| G | 0.072 | 0.070 | 0.063 | 0.018 | 0.005 | 0.073 | 0.001 | 0.000 | 1.000 | 0.310 | 0.159 | 0.191 |
| T | 0.409 | 0.422 | 0.574 | 0.896 | 0.971 | 0.700 | 0.135 | 0.000 | 0.000 | 0.124 | 0.361 | 0.276 |

# Weight Matrix – 3’ Splice Sites

```
SITE FREQUENCIES:
A  0.400  0.429  0.282  0.058  0.008  0.092  0.029  1.000  0.000  0.410  0.293  0.307
C  0.118  0.079  0.081  0.029  0.016  0.135  0.834  0.000  0.000  0.156  0.187  0.225
G  0.072  0.070  0.063  0.018  0.005  0.073  0.001  0.000  1.000  0.310  0.159  0.191
T  0.409  0.422  0.574  0.896  0.971  0.700  0.135  0.000  0.000  0.124  0.361  0.276

BACKGROUND FREQUENCIES:
A  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321
C  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179
G  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179  0.179
T  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321  0.321



WEIGHTS:
A   0.32   0.42  -0.18  -2.46  -5.29  -1.79  -3.45   1.64 -99.00   0.36  -0.13  -0.06
C  -0.60  -1.18  -1.15  -2.64  -3.51  -0.41   2.22 -99.00 -99.00  -0.20   0.06   0.33
G  -1.31  -1.35  -1.51  -3.35  -5.23  -1.30  -6.93 -99.00   2.48   0.79  -0.17   0.10
T   0.35   0.39   0.84   1.48   1.60   1.12  -1.24 -99.00 -99.00  -1.37   0.17  -0.22
```

# Scoring a Candidate 3’ Splice Site

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 0.32 | 0.42 | -0.18 | -2.46 | -5.29 | -1.79 | -3.45 | 1.64 | -99.00 | 0.36 | -0.13 | -0.06 |
| C | -0.60 | -1.18 | -1.15 | -2.64 | -3.51 | -0.41 | 2.22 | -99.00 | -99.00 | -0.20 | 0.06 | 0.33 |
| G | -1.31 | -1.35 | -1.51 | -3.35 | -5.23 | -1.30 | -6.93 | -99.00 | 2.48 | 0.79 | -0.17 | 0.10 |
| T | 0.35 | 0.39 | 0.84 | 1.48 | 1.60 | 1.12 | -1.24 | -99.00 | -99.00 | -1.37 | 0.17 | -0.22 |

**T T C T T A C A G A A T**

0.35 + 0.39 +-1.15 + 1.48 + 1.60 +-1.79 + 2.22 + 1.64 + 2.48 + 0.36 +-0.13 +-0.22 = 7.23

- General def.: a *weight matrix* $W$ has

  entries $w_{rj}$ indexed by residues $r \in A$, and $1 \leq j \leq n$

- *score* of a sequence $s = (s_1\, s_2 \ldots s_n)$ is

$$\sum_{1 \leq j \leq n} w_{s_j j}$$

- In the site case,

$$w_{rj} = \log(P_j(r \mid M_{\text{site}})) - \log(P_j(r \mid M_{\text{background}}))$$

# Simple Hypothesis Testing

- Suppose we wish to decide between two models:
  - $M_a$ (the *alternative hypothesis*), and
  - $M_0$ (the *null hypothesis*)

  using an observation $s$ from a sample space $S$. (e.g.
  - $s$ a sequence,
  - $M_a$ a site model
  - $M_0$ a "background" (non-site) model.
- Strategy:
  - choose a subset $C \subset S$, called the *critical region* for the comparison.
  - If $s$ falls within $C$, reject $M_0$ (accept $M_a$),
  - otherwise accept $M_0$ (reject $M_a$).

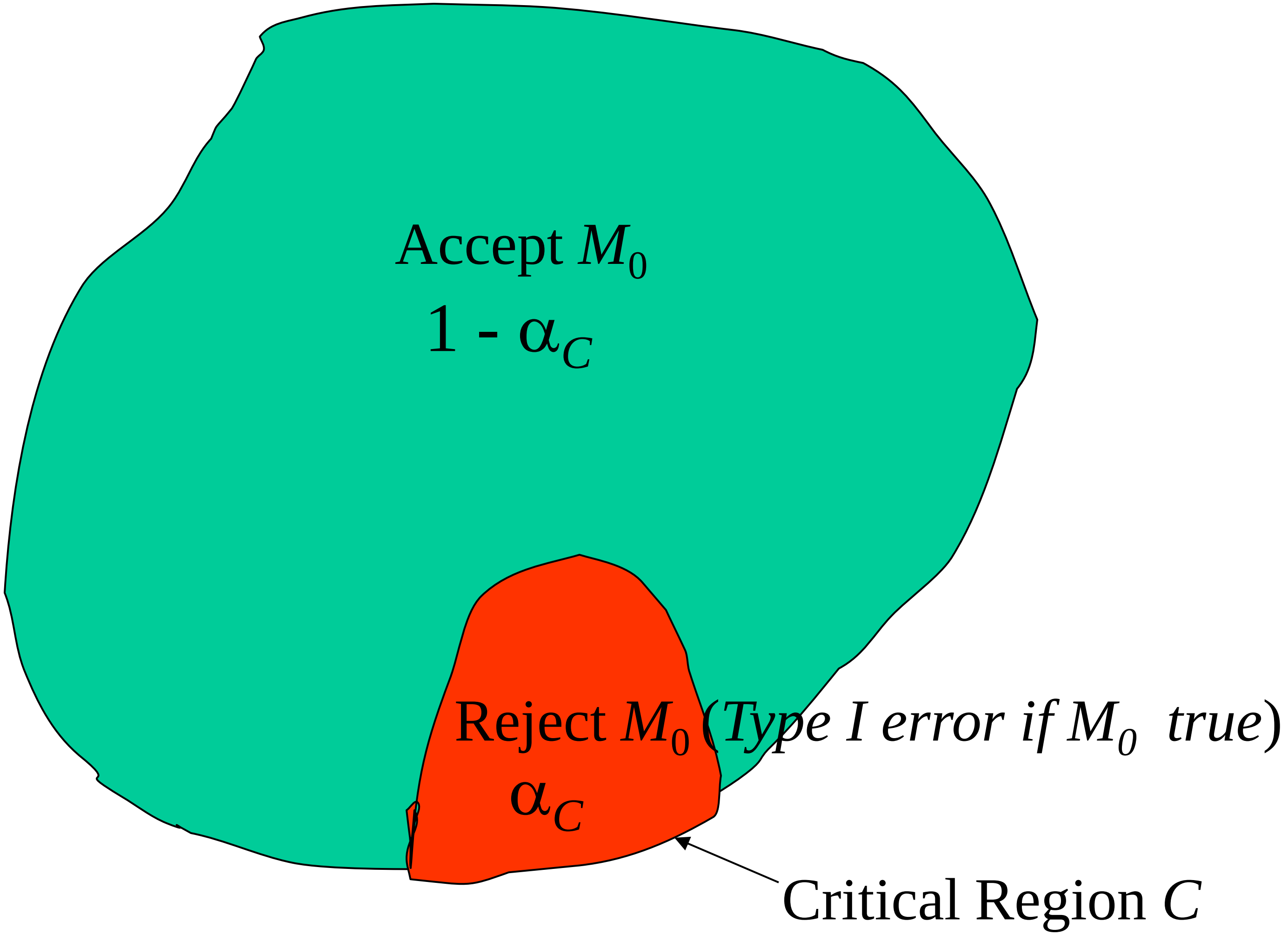
# Types of Errors with Hypothesis Test

- a *Type I error* occurs if we reject $M_0$ when it is true.
  - For a given critical region $C$, the prob of committing a Type I error is denoted $\alpha_C$

$$\alpha_C = P(C \mid M_0) = \Sigma_{s \in C} P(s \mid M_0)$$

- $\alpha_C$ is called the *significance level* of the test

# Sample Space *S* – probabilities under $M_0$

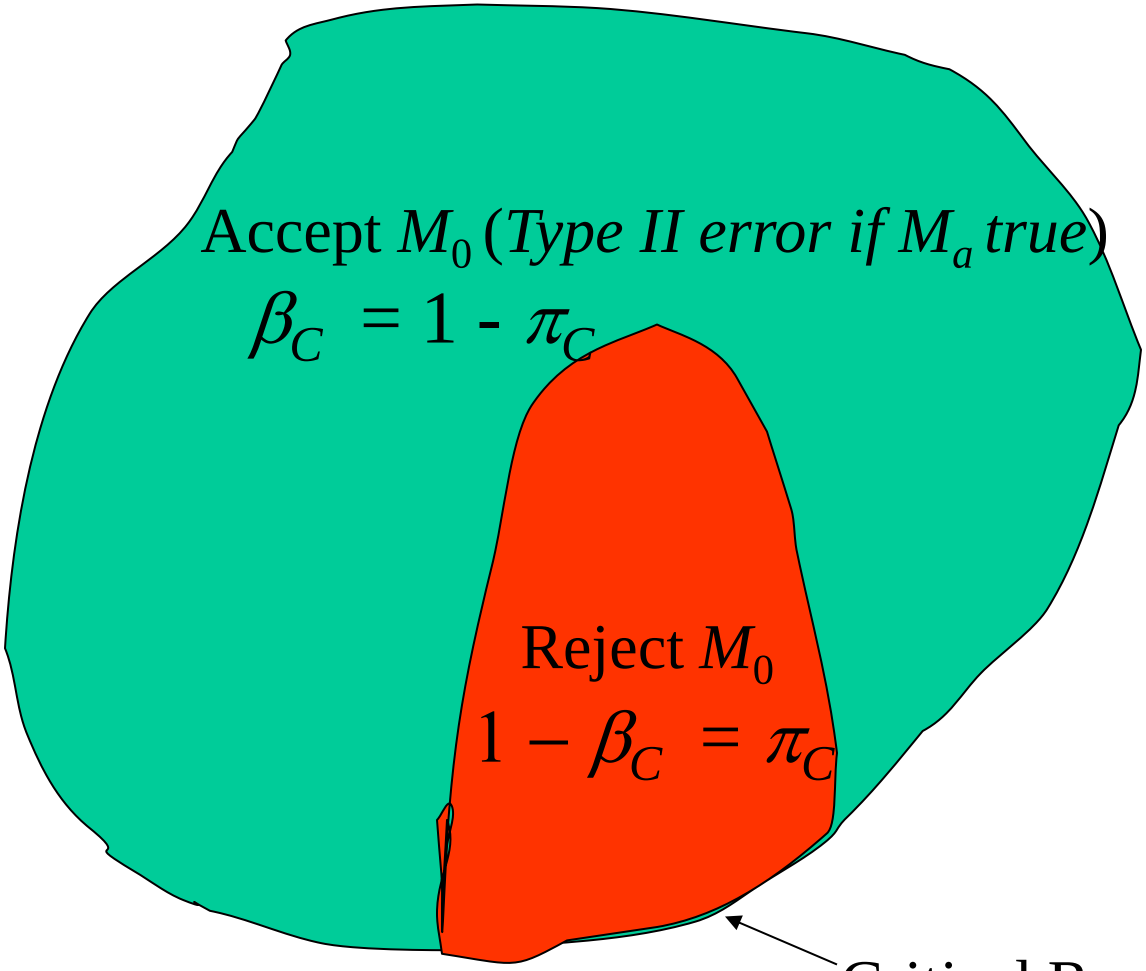


- a *Type II error* occurs if we accept $M_0$ when it is false.
  - For a given $C$, prob of committing a Type II error is denoted $\beta_C$

$$\beta_C = \Sigma_{s \notin C} P(s \mid M_a) = 1 - P(C \mid M_a)$$

- $\pi_C = 1 - \beta_C$ is called the *power* of the test.

# Sample Space $S$ – probabilities under $M_a$



Critical Region $C$

- Designing a test involves a tradeoff between significance and power
  - smaller *C* gives smaller Type I error but larger Type II error (lower power).

# Likelihood Ratio Tests

- A *likelihood ratio test* of models $M_a$ and $M_0$ is a hypothesis test of the two models, with critical region $C$ defined by

$$C = C_\Lambda = \{s \mid LR(M_a, M_0 \mid s) \geq \Lambda\}$$

for some non-negative constant $\Lambda$, the *cutoff value*.

- Neyman-Pearson lemma motivates use of the *likelihood ratio* as an optimal *discriminator*, or "score"
  - even in contexts where we aren't explicitly testing hypotheses.
- any monotonic function $f(LR)$ of likelihood ratio has equivalent optimality properties
  - because defines the same set of critical regions:

$$LR(M_a, M_0 \mid s) \geq \Lambda \Leftrightarrow f(LR(M_a, M_0 \mid s)) \geq f(\Lambda)$$

- convenient to take $f$ to be the log function, in which case we get the *log likelihood ratio*.

# Neyman-Pearson lemma

Let $M_a$ and $M_0$ be two models, and $C_\Lambda$ the critical region defined by a likelihood ratio test of $M_a$ vs. $M_0$ with

- cutoff value $\Lambda$,
- significance level $\alpha_\Lambda$, and
- power $\pi_\Lambda = 1 - \beta_\Lambda$.

*Then* if $C$ is any other critical region, we have

- If $\alpha_C < \alpha_\Lambda$ , then $\pi_C < \pi_\Lambda$ (and $\beta_C > \beta_\Lambda$)
- If $\alpha_C = \alpha_\Lambda$ , then $\pi_C \leq \pi_\Lambda$ (and $\beta_C \geq \beta_\Lambda$)

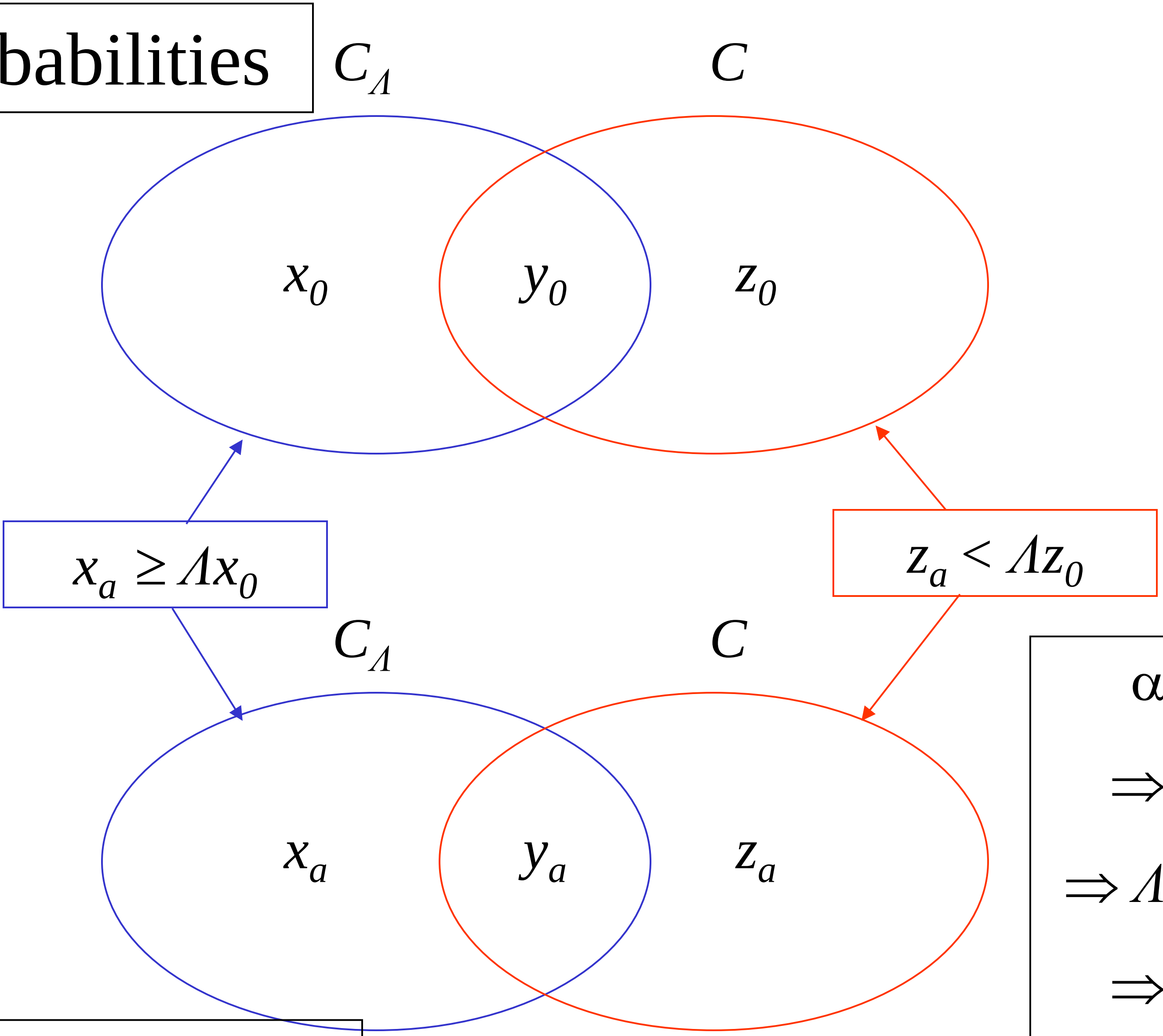
In other words, the likelihood ratio test with significance level $\alpha_\Lambda$ is the most powerful test

- (has the lowest type II error rate)

with that significance level.

# Idea of Neyman-Pearson lemma *proof*:

$M_0$ probabilities



$M_a$ probabilities

$$\alpha_C < \alpha_\Lambda$$

$$\Rightarrow z_0 < x_0$$

$$\Rightarrow \Lambda\, z_0 < \Lambda\, x_0$$

$$\Rightarrow z_a < x_a$$

$$\Rightarrow \pi_C < \pi_\Lambda$$

- ***Proof***: Suppose $\alpha_C < \alpha_\Lambda$ .Then

$$\sum_{s\in C} P(s \mid M_0) < \sum_{s\in C_\Lambda} P(s \mid M_0)$$

Subtract from both sides the terms involving $s \in C \cap C_\Lambda$ This leaves

$$(1) \quad \sum_{s\in C\setminus C_\Lambda} P(s \mid M_0) < \sum_{s\in C_\Lambda \setminus C} P(s \mid M_0)$$

- By definition of the likelihood ratio test, for any observation *s*,

$$s \in C_\Lambda \Leftrightarrow P(s \mid M_a) \geq \Lambda P(s \mid M_0)$$

- From this, it follows that

(2) $$\sum_{s \in C \backslash C_\Lambda} \frac{1}{\Lambda} P(s \mid M_a) < \sum_{s \in C \backslash C_\Lambda} P(s \mid M_0)$$

and

(3) $$\sum_{s \in C_\Lambda \backslash C} P(s \mid M_0) \leq \sum_{s \in C_\Lambda \backslash C} \frac{1}{\Lambda} P(s \mid M_a)$$

- Combining (2), (1), and (3)

$$\sum_{s\in C\backslash C_\Lambda} \frac{1}{\Lambda} P(s\,|\,M_a) < \sum_{s\in C\backslash C_\Lambda} P(s\,|\,M_0) < \sum_{s\in C_\Lambda\backslash C} P(s\,|\,M_0) \leq \sum_{s\in C_\Lambda\backslash C} \frac{1}{\Lambda} P(s\,|\,M_a)$$

so (cancelling the common factor 1 / $\Lambda$)

$$\sum_{s\in C\backslash C_\Lambda} P(s\,|\,M_a) < \sum_{s\in C_\Lambda\backslash C} P(s\,|\,M_a)$$

so, adding in the terms corresponding to $s \in C \cap C_\Lambda$

$$\sum_{s\in C} P(s\,|\,M_a) < \sum_{s\in C_\Lambda} P(s\,|\,M_a)$$

i.e $\pi_C < \pi_\Lambda$ The other part of the lemma ($\pi_C \leq \pi_\Lambda$ if $\alpha_C = \alpha_\Lambda$) is proved similarly.