Lecture 2: Probability
### Statistical Paradigms

<table>
<thead>
<tr>
<th></th>
<th>Statistical Estimator</th>
<th>Method of Estimation</th>
<th>Output</th>
<th>Data Complexity</th>
<th>Prior Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical</td>
<td>Cost Function</td>
<td>Analytical Solution</td>
<td>Point Estimate</td>
<td>Simple</td>
<td>No</td>
</tr>
<tr>
<td>Maximum Likelihood</td>
<td>Probability Theory</td>
<td>Numerical Optimization</td>
<td>Point Estimate</td>
<td>Intermediate</td>
<td>No</td>
</tr>
<tr>
<td>Bayesian</td>
<td>Probability Theory</td>
<td>Sampling</td>
<td>Probability Distribution</td>
<td>Complex</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The unifying principal for this course is statistical estimation based on **probability**.
Overview

- Basic probability
  - Joint, marginal, conditional probability
  - Bayes Rule
- Random variables
- Probability distribution
  - Discrete
  - Continuous
- Moments

One could spend ½ a semester on this alone...
Example

White-breasted fruit dove (*Ptilinopus rivoli*)

Yellow-bibbed fruit dove (*Ptilinopus solomonensis*)
<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
</tr>
</tbody>
</table>

$Pr(A) = \text{probability that event A occurs}$

$Pr(R) = \, ?$

$Pr(Rc) = \, ?$

$Pr(S) = \, ?$

$Pr(Sc) = \, ?$
<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
</tr>
</tbody>
</table>

Pr(A) = probability that event A occurs

Pr(R) = 11/32
Pr(Rc) = 21/32

Pr(S) = 20/32
Pr(Sc) = 12/32
Joint Probability

Pr(A,B) = probability that both A and B occur

Pr(R,Sc) = ?
Pr(S,Rc) = ?
Pr(R,S) = ?
Pr(Rc,Sc) = ?

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th>Pr(R) = 11/32</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>Pr(Rc) = 21/32</td>
</tr>
<tr>
<td></td>
<td>Pr(S) = 20/32</td>
<td>Pr(Sc) = 12/32</td>
<td>32</td>
</tr>
</tbody>
</table>
### Joint Probability

\[
Pr(A,B) = \text{probability that both } A \text{ and } B \text{ occur}
\]

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>Pr(R) = 11/32</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>Pr(Rc) = 21/32</td>
</tr>
</tbody>
</table>

\[
Pr(S) = 20/32 \quad Pr(Sc) = 12/32 \quad 32
\]

Pr(R,Sc) = 9/32
Pr(S,Rc) = 18/32
Pr(R,S) = 2/32
Pr(Rc,Sc) = 3/32
\[
\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A,B)
\]
\[
= 1 - \Pr(\text{neither})
\]

\[
\Pr(R \text{ or } S) = ?
\]
\[
\text{Pr}(A \text{ or } B) = \text{Pr}(A) + \text{Pr}(B) - \text{Pr}(A,B) \\
= 1 - \text{Pr}\text{(neither)}
\]

\[
\text{Pr}(R \text{ or } S) = \frac{11}{32} + \frac{20}{32} - \frac{2}{32} = \frac{29}{32} \\
= \frac{32}{32} - \frac{3}{32} = \frac{29}{32}
\]

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>Pr(R) = \frac{11}{32}</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>Pr(Rc) = \frac{21}{32}</td>
</tr>
<tr>
<td></td>
<td>Pr(S) = \frac{20}{32}</td>
<td>Pr(Sc) = \frac{12}{32}</td>
<td></td>
</tr>
</tbody>
</table>
If $\Pr(A,B) = \Pr(A) \cdot \Pr(B)$ then $A$ and $B$ are independent

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th>Pr(R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>$\frac{11}{32}$</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>$\frac{21}{32}$</td>
</tr>
</tbody>
</table>

| Pr(S) = $\frac{20}{32}$ | Pr(Sc) = $\frac{12}{32}$ | 32 |

If $\Pr(A,B) = \Pr(A) \cdot \Pr(B)$ then $A$ and $B$ are independent

$\Pr(R,S) = \Pr(R) \cdot \Pr(S)$ ??
If \( \Pr(A,B) = \Pr(A) \cdot \Pr(B) \) then \( A \) and \( B \) are **independent**

\[
0.0625 = \frac{2}{32} = \Pr(R,S) \neq \Pr(R) \cdot \Pr(S) = \frac{11}{32} \cdot \frac{20}{32} = 0.215
\]
Conditional Probability

\[ \Pr(A \mid B) = \text{Probability of } A \text{ given } B \text{ occurred} \]

\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]
\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]

\[ \Pr(R \mid S) = ? \]
### Conditional Probability

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Pr(S)</td>
<td>20/32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pr(Sc)</td>
<td>12/32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
Pr(B \mid A) = \frac{Pr(B,A)}{Pr(A)} \\
Pr(A \mid B) = \frac{Pr(A,B)}{Pr(B)} \\
Pr(R \mid S) = \frac{Pr(R,S)}{Pr(S)} = \frac{2/32}{20/32} = 2/20
\]
### Conditional Probability

\[
\Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)}
\]
\[
\Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)}
\]

\[
\Pr(S \mid R) = \frac{\Pr(S,R)}{\Pr(R)} = \frac{2/32}{11/32} = 2/11
\]
Joint = Conditional \cdot Marginal

\[ \Pr(B \mid A) = \frac{\Pr(B, A)}{\Pr(A)} \]

\[ \Pr(B, A) = \Pr(B \mid A) \cdot \Pr(A) \]

\[ \Pr(S, R) = \Pr(S \mid R) \cdot \Pr(R) \]

\[ \frac{2}{32} = \frac{2}{11} \cdot \frac{11}{32} \]
### Events

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th>Pr</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>$\frac{11}{32}$</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>$\frac{21}{32}$</td>
</tr>
</tbody>
</table>

| Pr(S) = $\frac{20}{32}$ | Pr(Sc) = $\frac{12}{32}$ | 32 |

### Marginal Probability

$$\Pr(B) = \sum \Pr(B, A_i)$$

$$\Pr(R) = ?$$
### Marginal Probability

\[
\Pr(B) = \sum \Pr(B, A_i)
\]

\[
\Pr(R) = \Pr(R, S) + \Pr(R, Sc)
\]

\[
= \frac{2}{32} + \frac{9}{32}
\]

\[
= \frac{11}{32}
\]

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th>Marginal</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>(\Pr(R) = \frac{11}{32})</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>(\Pr(Rc) = \frac{21}{32})</td>
</tr>
<tr>
<td>Marginal</td>
<td>(\Pr(S) = \frac{20}{32})</td>
<td>(\Pr(Sc) = \frac{12}{32})</td>
<td>32</td>
</tr>
</tbody>
</table>
Marginal Probability

\[
\Pr(B) = \sum \Pr(B, A_i)
\]

\[
\Pr(B) = \sum \Pr(B | A_i) \cdot \Pr(A_i)
\]

\[
\Pr(R) = ?
\]
**Marginal Probability**

\[ \Pr(B) = \sum \Pr(B \mid A_i) \cdot \Pr(A_i) \]

\[ \Pr(R) = \Pr(R \mid S) \cdot \Pr(S) + \Pr(R \mid Sc) \cdot \Pr(Sc) \]

\[ = \frac{2}{20} \cdot \frac{20}{32} + \frac{9}{12} \cdot \frac{12}{32} \]

\[ = \frac{2}{32} + \frac{9}{32} \]

\[ = \frac{11}{32} \]
Conditional Probability

\[ \Pr(B | A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A | B) = \frac{\Pr(A,B)}{\Pr(B)} \]
Conditional Probability

\[ Pr(B \mid A) = \frac{Pr(B,A)}{Pr(A)} \]
\[ Pr(A \mid B) = \frac{Pr(A,B)}{Pr(B)} \]

\[ Pr(A,B) = Pr(B \mid A) \cdot Pr(A) \]
\[ Pr(B,A) = Pr(A \mid B) \cdot Pr(B) \]
Conditional Probability

\[
\Pr(B \mid A) = \frac{\Pr(B, A)}{\Pr(A)}
\]

\[
\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}
\]

\[
\Pr(A, B) = \Pr(B \mid A) \cdot \Pr(A)
\]

\[
\Pr(B, A) = \Pr(A \mid B) \cdot \Pr(B)
\]

Joint = Conditional x Marginal
Conditional Probability

\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]

\[ \Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A) \]
\[ \Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B) \]

**Joint = Conditional x Marginal**

**Competition: Best mnemonic**
Conditional Probability

\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]

\[ \Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A) \]
\[ \Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B) \]

\[ \Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A) \]
Conditional Probability

\[ \Pr(B \mid A) = \frac{\Pr(B,A)}{\Pr(A)} \]
\[ \Pr(A \mid B) = \frac{\Pr(A,B)}{\Pr(B)} \]

\[ \Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A) \]
\[ \Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B) \]

\[ \Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A) \]
\[ \Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)} \]

BAYES RULE
### Bayes Rule

\[ \text{Pr}(A \mid B) = \frac{\text{Pr}(B \mid A) \cdot \text{Pr}(A)}{\text{Pr}(B)} \]

\[ \text{Pr}(R \mid S) = \frac{\text{Pr}(S \mid R) \cdot \text{Pr}(R)}{\text{Pr}(S)} \]

\[ = \frac{(2/11) \cdot (11/32)}{(20/32)} \]

\[ = 2/20 \]

<table>
<thead>
<tr>
<th>Events</th>
<th>S</th>
<th>Sc</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>2</td>
<td>9</td>
<td>Pr(R) = 11/32</td>
</tr>
<tr>
<td>Rc</td>
<td>18</td>
<td>3</td>
<td>Pr(Rc) = 21/32</td>
</tr>
<tr>
<td></td>
<td>Pr(S) = 20/32</td>
<td>Pr(Sc) = 12/32</td>
<td>32</td>
</tr>
</tbody>
</table>
BAYES RULE: alternate form

\[ Pr(A \mid B) = Pr(B \mid A) \cdot Pr(A) / Pr(B) \]

\[ Pr(A \mid B) = Pr(B \mid A) \cdot Pr(A) / \sum Pr(B,A_i) \]

\[ Pr(A \mid B) = Pr(B \mid A) \cdot Pr(A) / \left( \sum Pr(B \mid A_i) \cdot Pr(A_i) \right) \]

Normalizing Constant
Monty Hall Problem
Monty Hall Problem

\[ P_1 = \frac{1}{3} \]
\[ P_2 = \frac{1}{3} \]
\[ P_3 = \frac{1}{3} \]
Monty Hall Problem

\[ P_1 \mid \text{open 3} = ? \]
\[ P_2 \mid \text{open 3} = ? \]
\[ P_3 \mid \text{open 3} = ? \]
Monty Hall Problem

open 3 | P₁
open 3 | P₂
open 3 | P₃
Monty Hall Problem

open 3 | $P_1$ = 1/2

open 3 | $P_2$ = 1

open 3 | $P_3$ = 0
Monty Hall Problem

\[ P_1 \mid \text{open 3} = \frac{1}{3} \]
\[ P_2 \mid \text{open 3} = \frac{2}{3} \]
\[ P_3 \mid \text{open 3} = 0 \]
Monty Hall Problem

1. Player picks car (probability 1/3)
   - Host reveals either goat
   - Switching loses.

2. Player picks Goat A (probability 1/3)
   - Host must reveal Goat B
   - Switching wins.

3. Player picks Goat B (probability 1/3)
   - Host must reveal Goat A
   - Switching wins.
\[ P(I'm \text{ near} \mid I \text{ picked up} \text{ a seashell}) = \]
\[ \frac{P(I \text{ picked up} \mid I'm \text{ near}) P(I \text{ picked up} \text{ a seashell})}{P(I'm \text{ near})} \]

Statistically speaking, if you pick up a seashell and don't hold it to your ear, you can probably hear the ocean.
BAYES RULE: alternate form

\[ \Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B) \]

\[ \Pr(A\mid B) = \Pr(B\mid A)\cdot\Pr(A) / (\sum \Pr(B\mid A)\cdot\Pr(A) ) \]

Factoring probabilities:

\[ \Pr(A, B, C) = \Pr(A\mid B, C)\Pr(B, C) \]

\[ = \Pr(A\mid B, C)\Pr(B\mid C)\Pr(C) \]
BAYES RULE: alternate form

\[ \Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B) \]

\[ \Pr(A\mid B) = \Pr(B\mid A) \cdot \Pr(A) / (\sum \Pr(B\mid A) \cdot \Pr(A)) \]

Factoring probabilities:

\[ \Pr(A,B,C) = \Pr(A\mid B,C) \Pr(B,C) \]

\[ = \Pr(A\mid B,C) \Pr(B\mid C) \Pr(C) \]

Joint = Conditional x Marginal
Random Variables

“a variable that can take on more than one value, in which the values are determined by probabilities”

\[ \Pr(Z = z_k) = p_k \]

given:

\[ 0 \leq p_k \leq 1 \]

Random variables can be continuous or discrete
Discrete random variables

$z_k$ can only take on discrete values (typically integers)

We can define two important and interrelated functions

**probability mass function (pmf):**

$$f(z) = Pr(Z = z_k) = p_k$$

where $\Sigma f(z) = 1$ (if not met, is just a density fcn)

**Cumulative distribution function (cdf):**

$$F(z) = Pr(Z \leq z_k) = \Sigma f(z) \text{ summed up to } k$$

$0 \leq F(z) \leq 1$ but can be infinite in $z$
Example

For the set \{1,2,3,4,5,6\}

\[ f(z) = \Pr(Z = z_k) = \{1/6,1/6,1/6,1/6,1/6,1/6\} \]

\[ F(z) = \Pr(Z \leq z_k) = \{1/6,2/6,3/6,4/6,5/6,6/6\} \]

For \( z < 1 \), \( f(z) = 0 \), \( F(z) = 0 \)
For \( z > 6 \), \( f(z) = 0 \), \( F(z) = 1 \)
Continuous Random Variables

$z$ is Real (though can still be bound)

**Cumulative distribution function (cdf):**

$$F(z) = \Pr(Z \leq z) \quad \text{where } 0 \leq F(z) \leq 1$$

$$\Pr(Z = z) \text{ is infinitely small}$$

$$\Pr(z \leq Z \leq z + dz) = \Pr(Z \leq z + dz) - \Pr(Z \leq z)$$

$$= F(z + dz) - F(z)$$

**Probability density function (pdf):**

$$f(z) = \frac{dF}{dz}$$

$$f(z) \geq 0 \text{ but NOT bound by 1}$$
Continuous Random Variables

\[ f \text{ is derivative of } F \]
\[ F \text{ is integral of } f \]

\[ Pr(z \leq Z \leq z + dz) = \int_{z}^{z+dz} f(z) \]

ANY function that meets these rules (positive, integrate to 1)

Wednesday we will be going over a number of standard number of distributions and discussing their interpretation/application.
Example: exponential distribution

\[ f(z) = \lambda \exp(-\lambda z) \]

\[ F(z) = 1 - \exp(-\lambda z) \]

Where \( z \geq 0 \)

What are the values of \( F(z) \) and \( f(z) \):
- At \( z = 0 \)?
- As \( z \to \infty \)?

What do \( F(z) \) and \( f(z) \) look like?
Exponential

\[ \text{Exp}(x|\lambda) = \lambda \exp(-\lambda x) \]

At \( z = 0 \)
- \( f(z) = 0 \)
- \( F(z) = 0 \)

As \( z \to \infty \)
- \( f(z) = 0 \)
- \( F(z) = 1 \)
Moments of probability distributions

\[ E[x^n] = \int x^n \cdot f(x) \, dx \]

E[ ] = Expected value

First moment (n=1) = mean

Example: exponential

\[ E[x] = \int x \cdot f(x) \, dx = \int_0^\infty x \lambda \exp(-\lambda x) \, dx \]

\[ = -x \exp(-\lambda x) \bigg|_0^\infty + \frac{1}{\lambda} \int_0^\infty \lambda \exp(-\lambda x) \, dx \]

\[ E[x] = 1/\lambda \]
Properties of means

\[ E[c] = c \]

\[ E[x + c] = E[x] + c \]

\[ E[cx] = c \ E[x] \]

\[ E[x+y] = E[x] + E[y] \]
  \hspace{1cm} \text{(even if } X \text{ is not independent of } Y) \]

\[ E(xy) = E[x]E[Y] \]
  \hspace{1cm} \text{only if independent} \]

\[ E[g(x)] \neq g(E[x]) \]
  \hspace{1cm} \text{Jensen's Inequality}
Central Moments

\[ E[(x - E[x])^n] = \int (x - E[x])^n \cdot f(x) \, dx \]

Second Central Moment = Variance = \( \sigma^2 \)
\[ \text{Var}(aX) = a^2 \text{Var}(X) \]

\[ \text{Var}(X + b) = \text{Var}(X) \]

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]

\[ \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab\text{Cov}(X, Y) \]

\[ \text{Var} \left( \sum X \right) = \sum \text{Var}(X_i) + 2 \sum_{i<j} \text{Cov}(X_i, X_j) \]

\[ \text{Var}(X) = \text{Var}(E[X|Y]) + E[\text{Var}(X|Y)] \]
Distributions and Probability

all the same properties apply to random variables

\[ Pr(A, B) \]  
joint distribution

\[ Pr(A|B) = \frac{Pr(A, B)}{Pr(B)} \]  
conditional distribution

\[ Pr(A) = \sum Pr(A|B_i) Pr(B_i) \]  
marginal distribution

\[ Pr(A|B) = \frac{Pr(B|A) Pr(A)}{Pr(B)} \]  
Baye's Rule
Looking forward...

Use probability distributions to:

- Quantify the match between models and data
- Represent uncertainty about model parameters
- Partition sources of process variability