Lecture 2: Probability

Statistical Paradigms

	Statistical Estimator	Method of Estimation	Output	Data Complexity	Prior Info
Classical	Cost Function	Analytical Solution	Point Estimate	Simple	No
Maximum Likelihood	Probability Theory	Numerical Optimization	Point Estimate	Intermediate	No
Bayesian	Probability Theory	Sampling	Probability Distribution	Complex	Yes

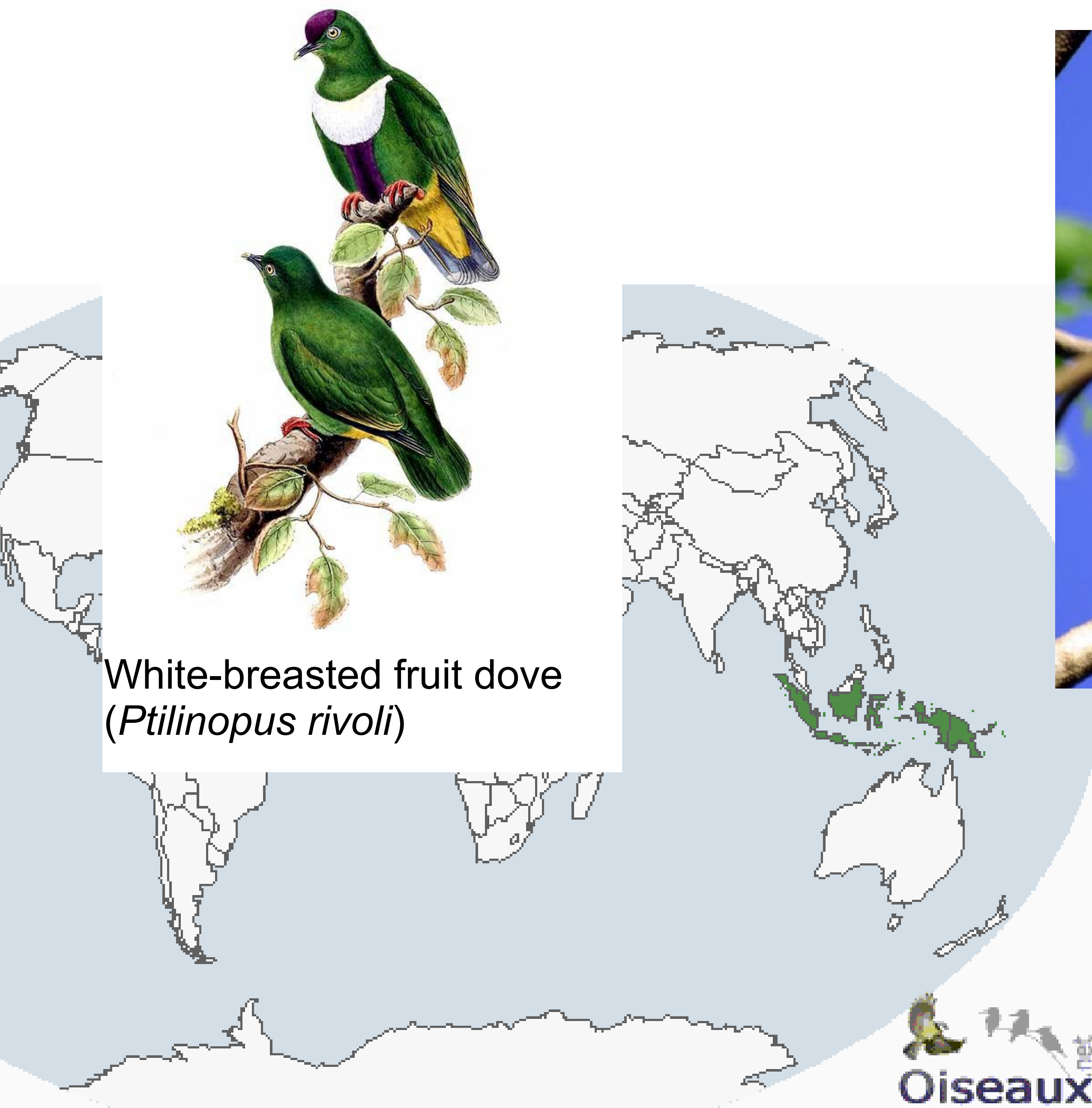
The unifying principal for this course is statistical estimation based on **probability**

Overview

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

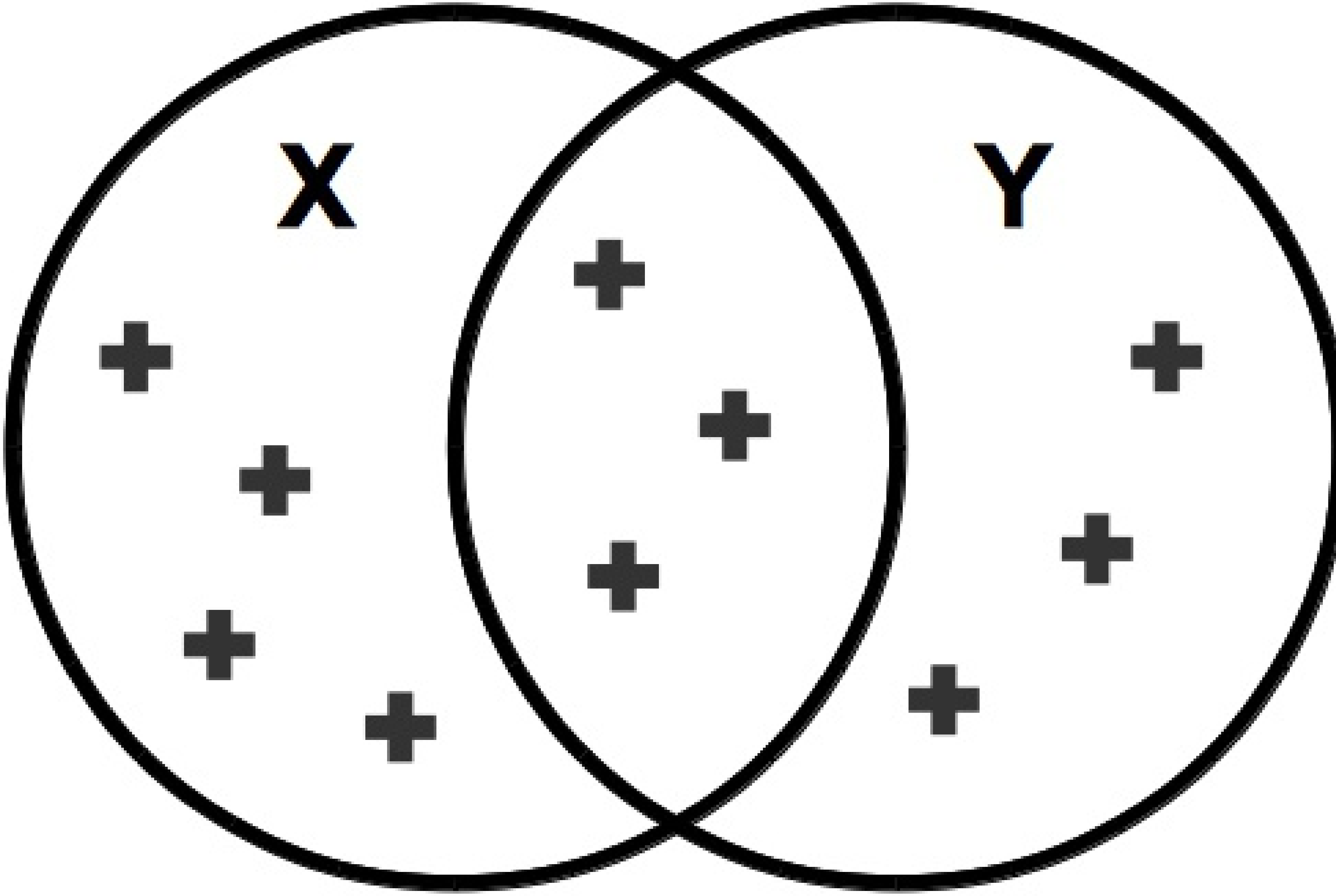
One could spend ½ a semester on this alone...

Example



White-breasted fruit dove
(*Ptilinopus rivoli*)

Yellow-bibbed fruit dove
(*Ptilinopus solomonensis*)


X
Y

Events	S	Sc	
R	2	9	
Rc	18	3	

Pr(A) = probability that event A occurs

Pr(R) = ?
Pr(Rc) = ?

Pr(S) = ?
Pr(Sc) = ?

Events	S	Sc	
R	2	9	
Rc	18	3	

Pr(A) = probability that event A occurs

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

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

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Joint Probability

Pr(A,B) = probability that both <u>A and B</u> occur

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

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Joint Probability

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

Pr(R,Sc) = 9/32
Pr(S,Rc) = 18/32
Pr(R,S) = 2/32
Pr(Rc,Sc) = 3/32

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Pr(A or B) = Pr(A) + Pr(B) – Pr(A,B)
= 1 – Pr(neither)

Pr(R or S) = ?

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Pr(A or B) = Pr(A) + Pr(B) – Pr(A,B)
= 1 – Pr(neither)

Pr(R or S) = 11/32 + 20/32 – 2/32 = 29/32
= 32/32 – 3/32 = 29/32

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

If Pr(A,B) = Pr(A)·Pr(B) then A and B are **independent**

Pr(R,S) = Pr(R)·Pr(S) ??

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

If Pr(A,B) = Pr(A)·Pr(B) then A and B are **<u>independent</u>**

$0.0625 = 2/32 = \Pr(R,S) \neq \Pr(R) \cdot \Pr(S) = 11/32 \cdot 20/32 = 0.215$

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Conditional Probability

Pr(A | B) = Probability of **A** ***given*** **B** occurred

Pr(A | B) = Pr(A,B) / Pr(B)
Pr(B | A) = Pr(B,A) / Pr(A)

Pr(R | S) = ?

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Conditional Probability

Pr(B | A) = Pr(B,A) / Pr(A)
Pr(A | B) = Pr(A,B) / Pr(B)

Pr(R | S) = Pr(R,S) / Pr(S)
= (2/32) / (20/32) = 2/20

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Conditional Probability

Pr(B | A) = Pr(B,A) / Pr(A)
Pr(A | B) = Pr(A,B) / Pr(B)

Pr(S | R) = Pr(S,R) / Pr(R)
= (2/32) / (11/32) = 2/11

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Joint = Conditional · Marginal

$$\Pr(B \mid A) = \Pr(B , A) / \Pr(A)$$

$$\Pr(B , A) = \Pr(B \mid A) \cdot \Pr(A)$$

$$\Pr(S , R) = \Pr(S \mid R) \cdot \Pr(R)$$
$$2/32 = (2/11) \cdot (11/32)$$

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Marginal Probability

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

Pr(R) = ?

Events	S	Sc	Marginal
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
Marginal	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Marginal Probability

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

$$\begin{aligned} \Pr(R) &= \Pr(R , S) + \Pr(R , Sc) \\ &= 2/32 + 9/32 \\ &= 11/32 \end{aligned}$$

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Marginal Probability

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

$$Pr(B) = \Sigma\ Pr(B \mid A_i) \cdot Pr(A_i)$$

$$Pr(R) = ?$$

Events	S	Sc	Marginal
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
Marginal	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Marginal Probability

$$\Pr(B) = \Sigma \Pr(B \mid A_i) \cdot \Pr(A_i)$$

$$\begin{aligned}\Pr(R) &= \Pr(R \mid S) \cdot \Pr(S) + \Pr(R \mid Sc) \cdot \Pr(Sc) \\ &= 2/20 \cdot 20/32 + 9/12 \cdot 12/32 \\ &= 2/32 + 9/32 \\ &= 11/32\end{aligned}$$

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

$$\Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B)$$

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

$$\Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B)$$

Joint = Conditional x Marginal

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

$$\Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B)$$

Joint = Conditional x Marginal

Competition: Best mnemonic

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

$$\Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B)$$

$$\Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A)$$

Conditional Probability

$$\Pr(B \mid A) = \Pr(B,A) / \Pr(A)$$
$$\Pr(A \mid B) = \Pr(A,B) / \Pr(B)$$

$$\Pr(A,B) = \Pr(B \mid A) \cdot \Pr(A)$$
$$\Pr(B,A) = \Pr(A \mid B) \cdot \Pr(B)$$

$$\Pr(A \mid B) \cdot \Pr(B) = \Pr(B \mid A) \cdot \Pr(A)$$

$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B)$$

BAYES RULE

Events	S	Sc	
R	2	9	Pr(R) = 11/32
Rc	18	3	Pr(Rc) = 21/32
	Pr(S) = 20/32	Pr(Sc) = 12/32	32

Bayes Rule

$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B)$$

$$\begin{aligned} \Pr(R \mid S) &= \Pr(S \mid R) \cdot \Pr(R) / \Pr(S) \\ &= (2/11) \cdot (11/32) / (20/32) \\ &= 2/20 \end{aligned}$$

BAYES RULE: alternate form

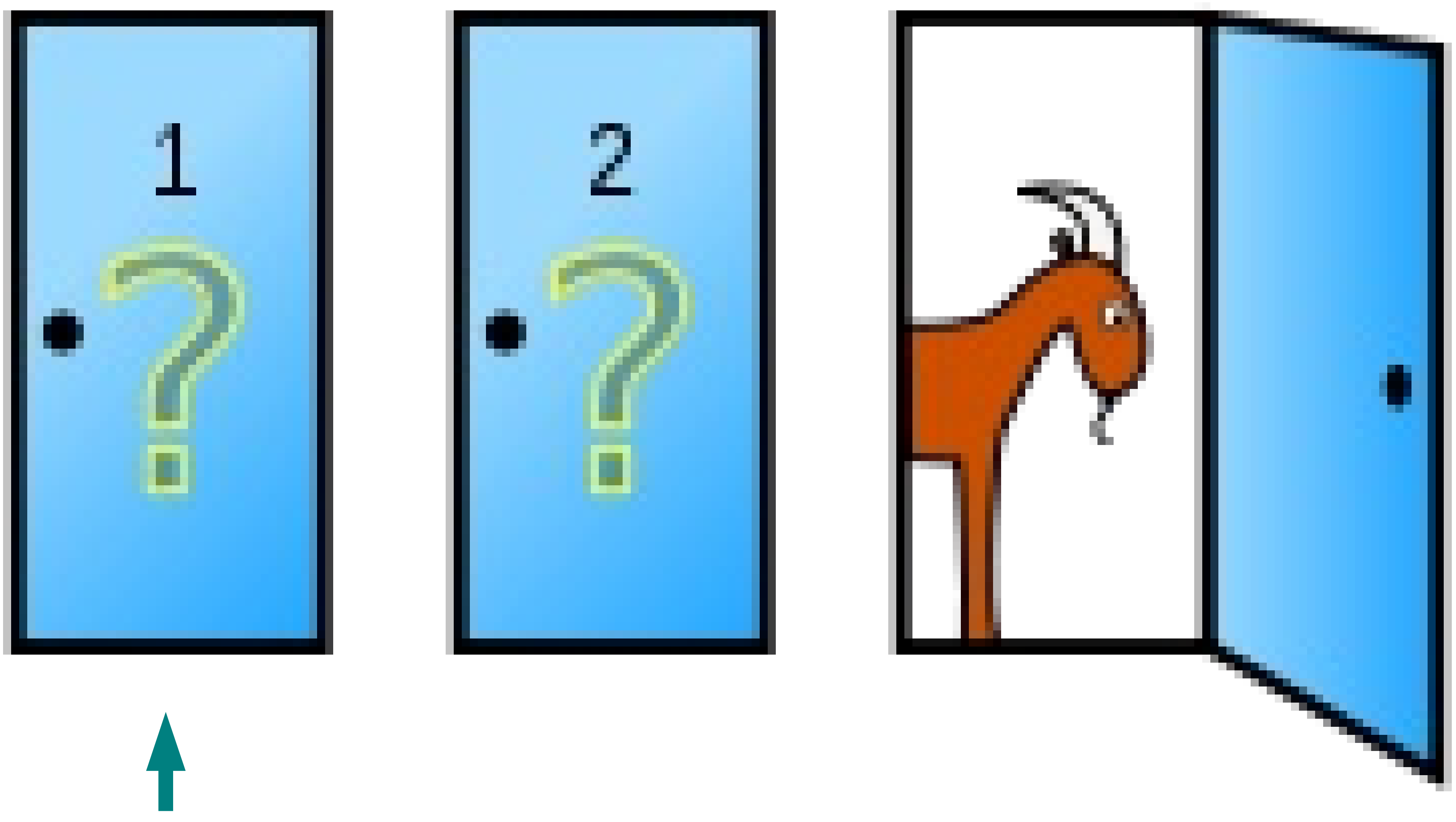
$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B)$$

$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Sigma \Pr(B, A_i)$$

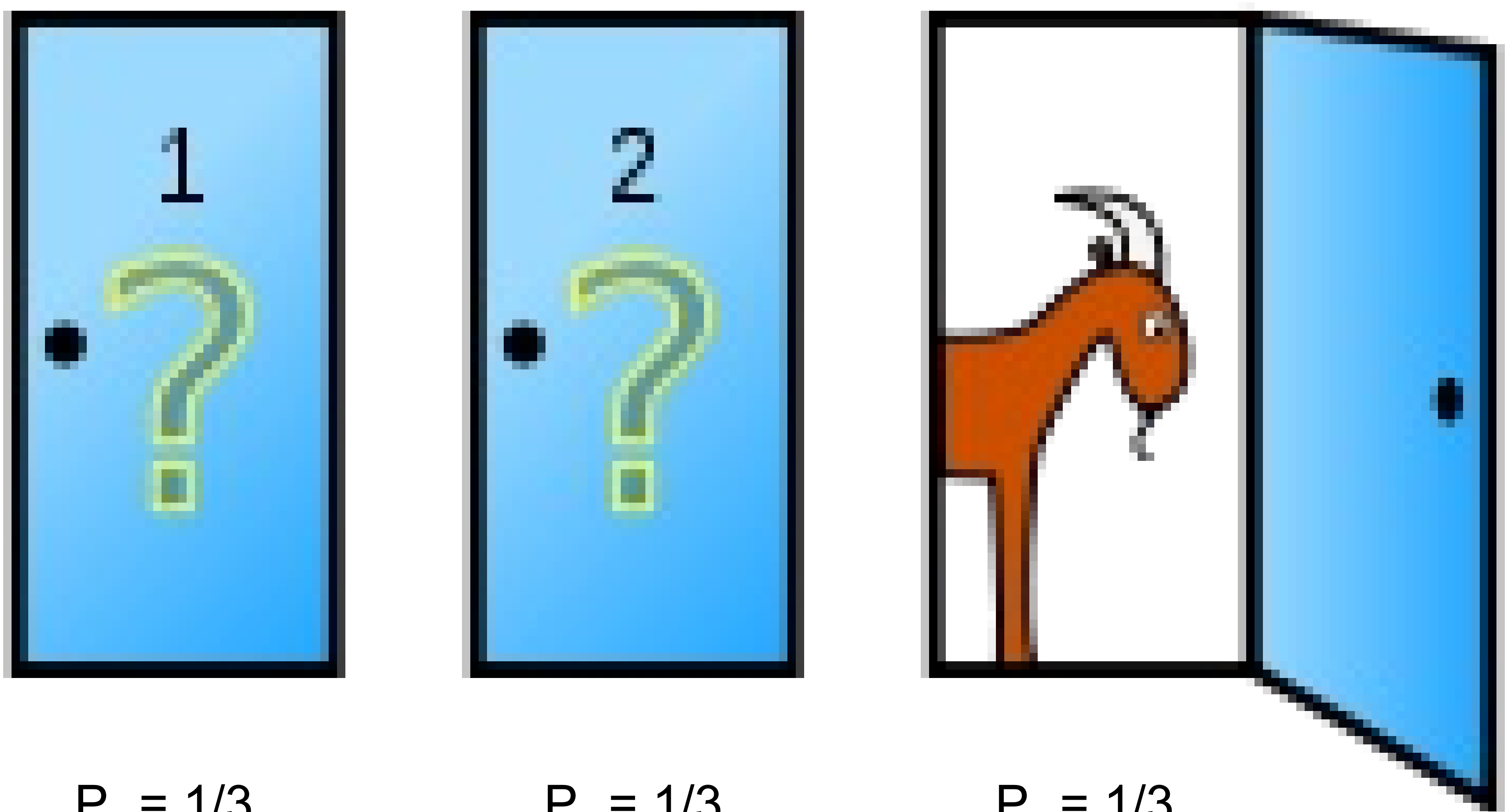
$$\Pr(A|B) = \Pr(B|A) \cdot \Pr(A) / \boxed{(\Sigma \Pr(B|A_i) \cdot \Pr(A_i))}$$

Normalizing Constant

Monty Hall Problem

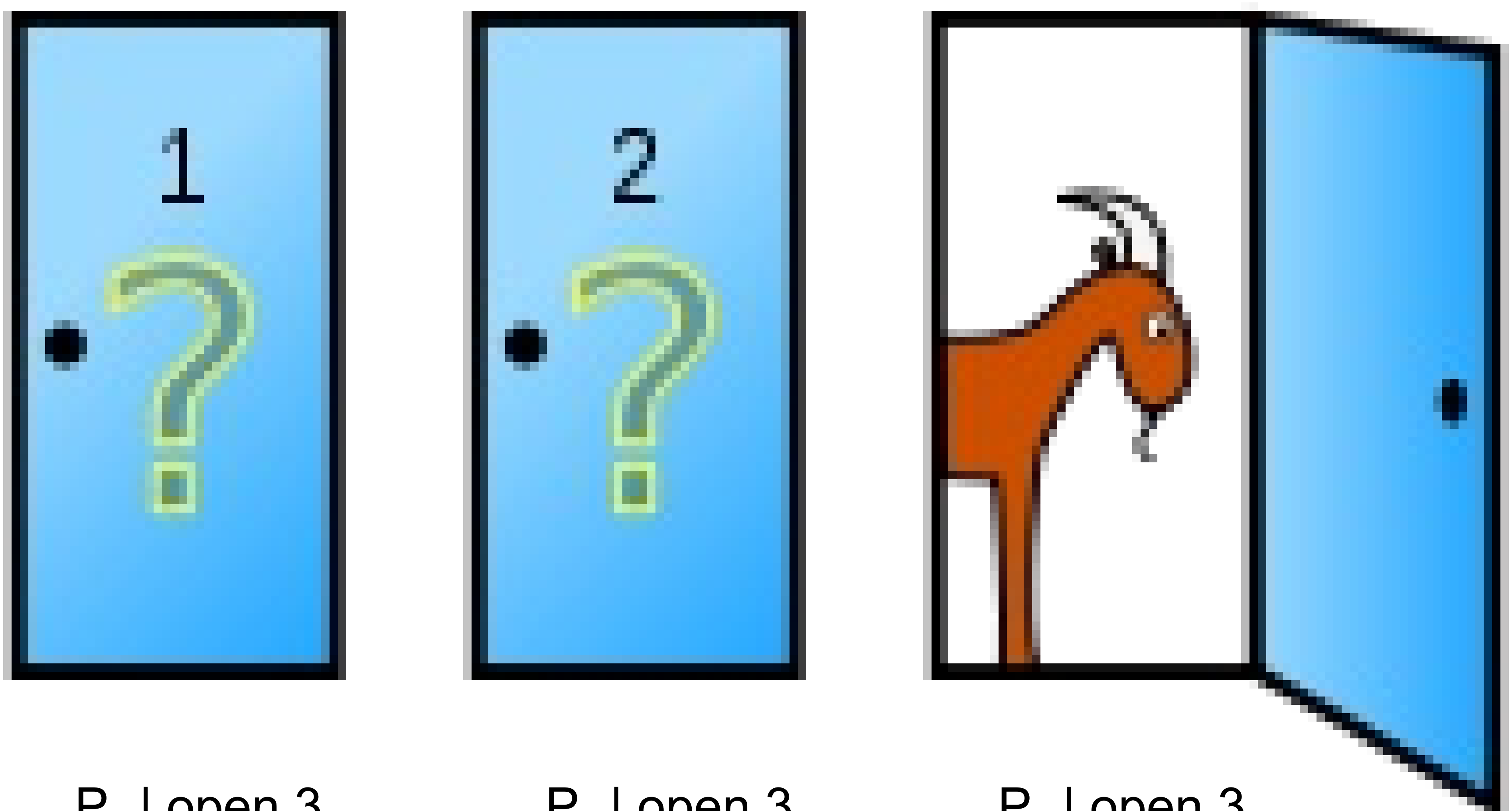


Monty Hall Problem



$P_1 = 1/3$ $P_2 = 1/3$ $P_3 = 1/3$

Monty Hall Problem

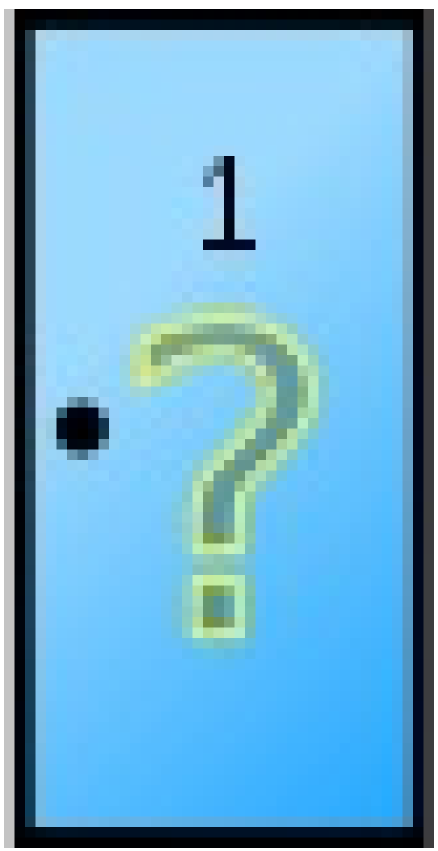


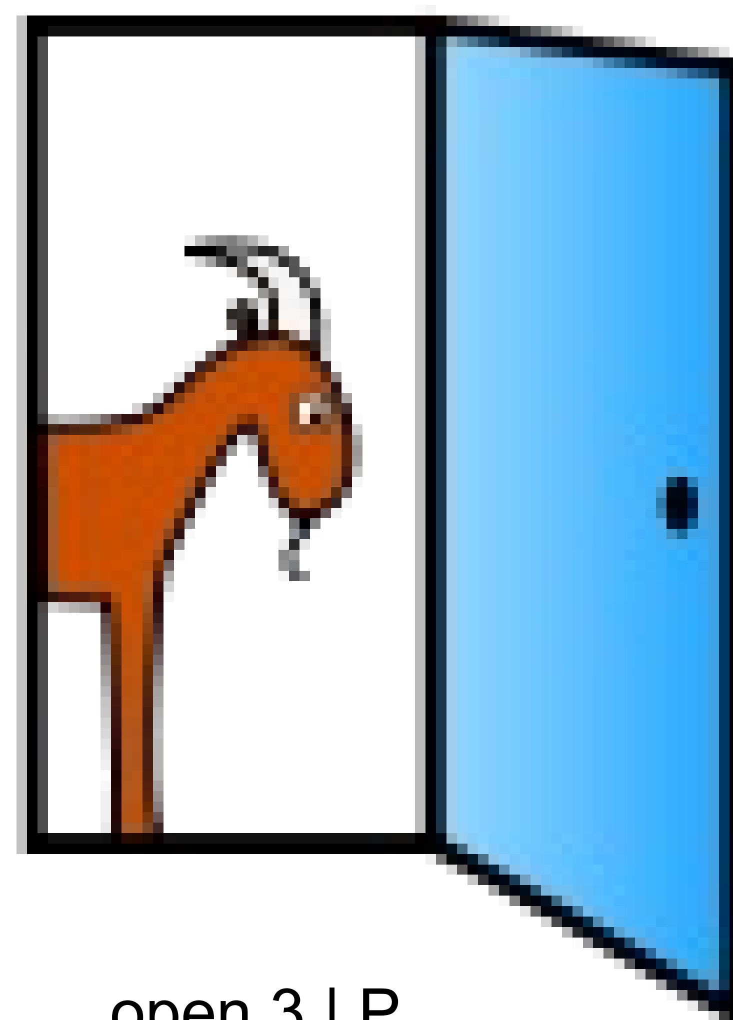
P_1 | open 3 = ?

P_2 | open 3 = ?

P_3 | open 3 = ?

Monty Hall Problem



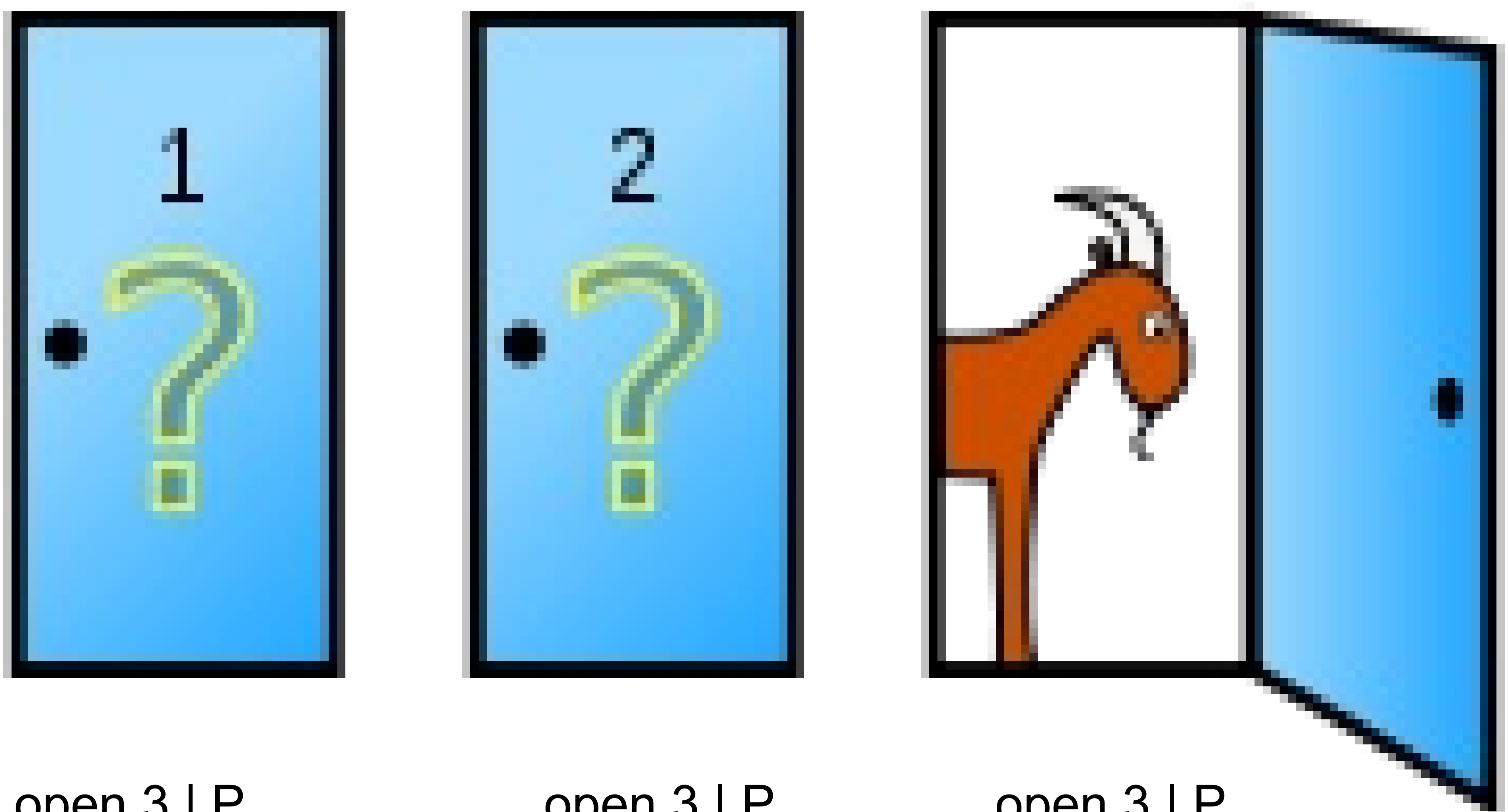


open 3 | P_1

open 3 | P_2

open 3 | P_3

Monty Hall Problem

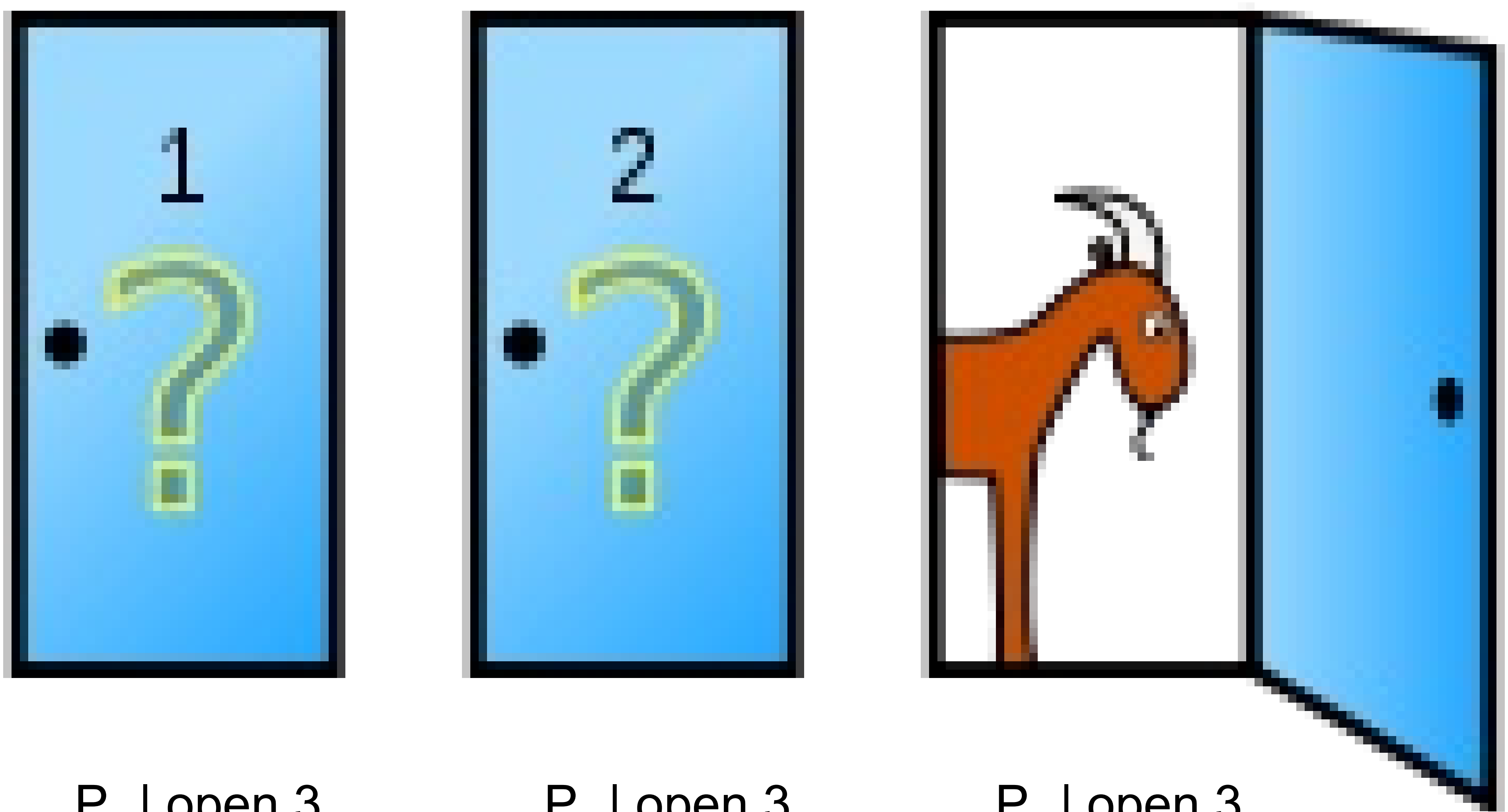


open 3 | P_1
= 1/2

open 3 | P_2
= 1

open 3 | P_3
= 0

Monty Hall Problem

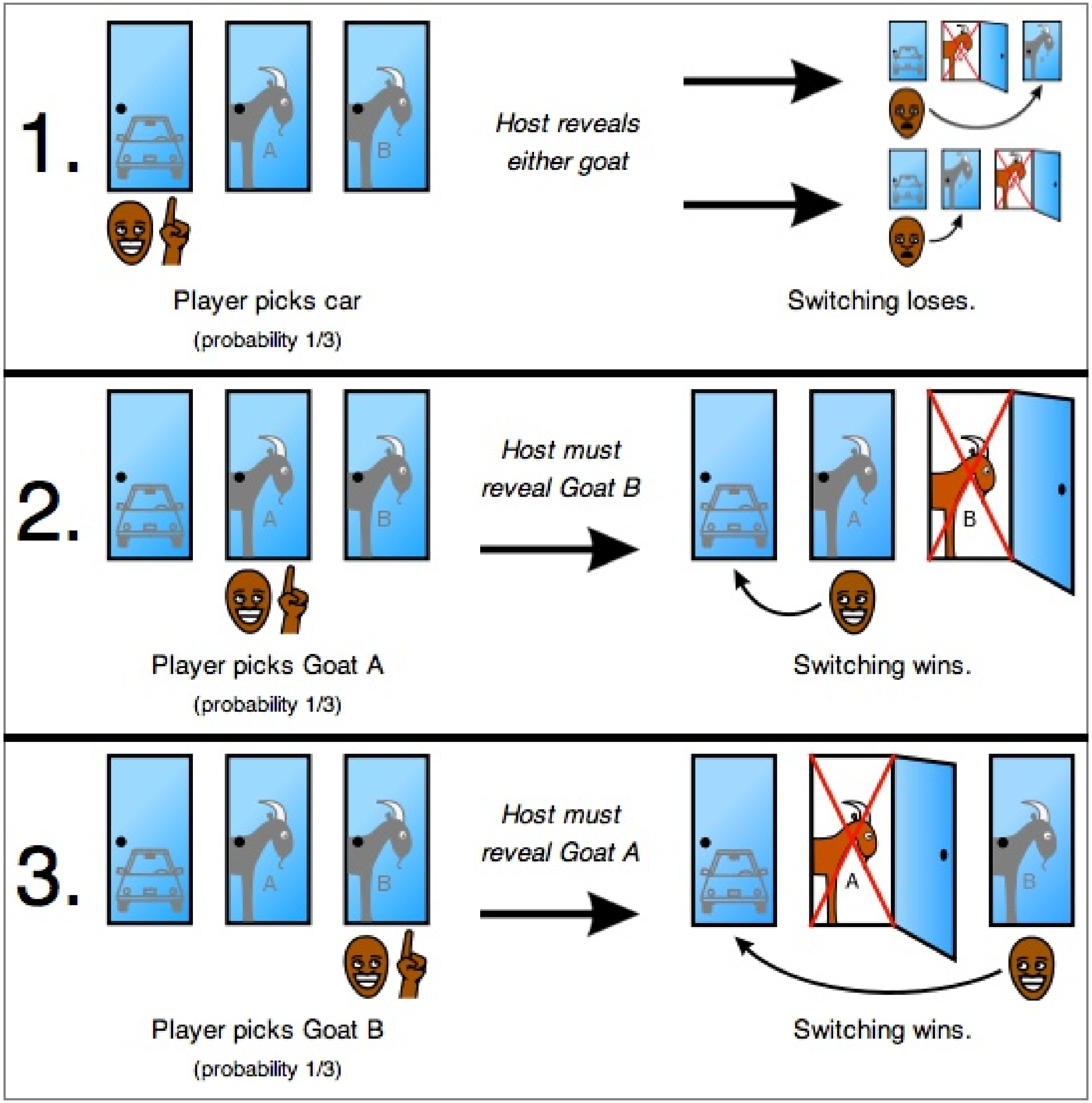


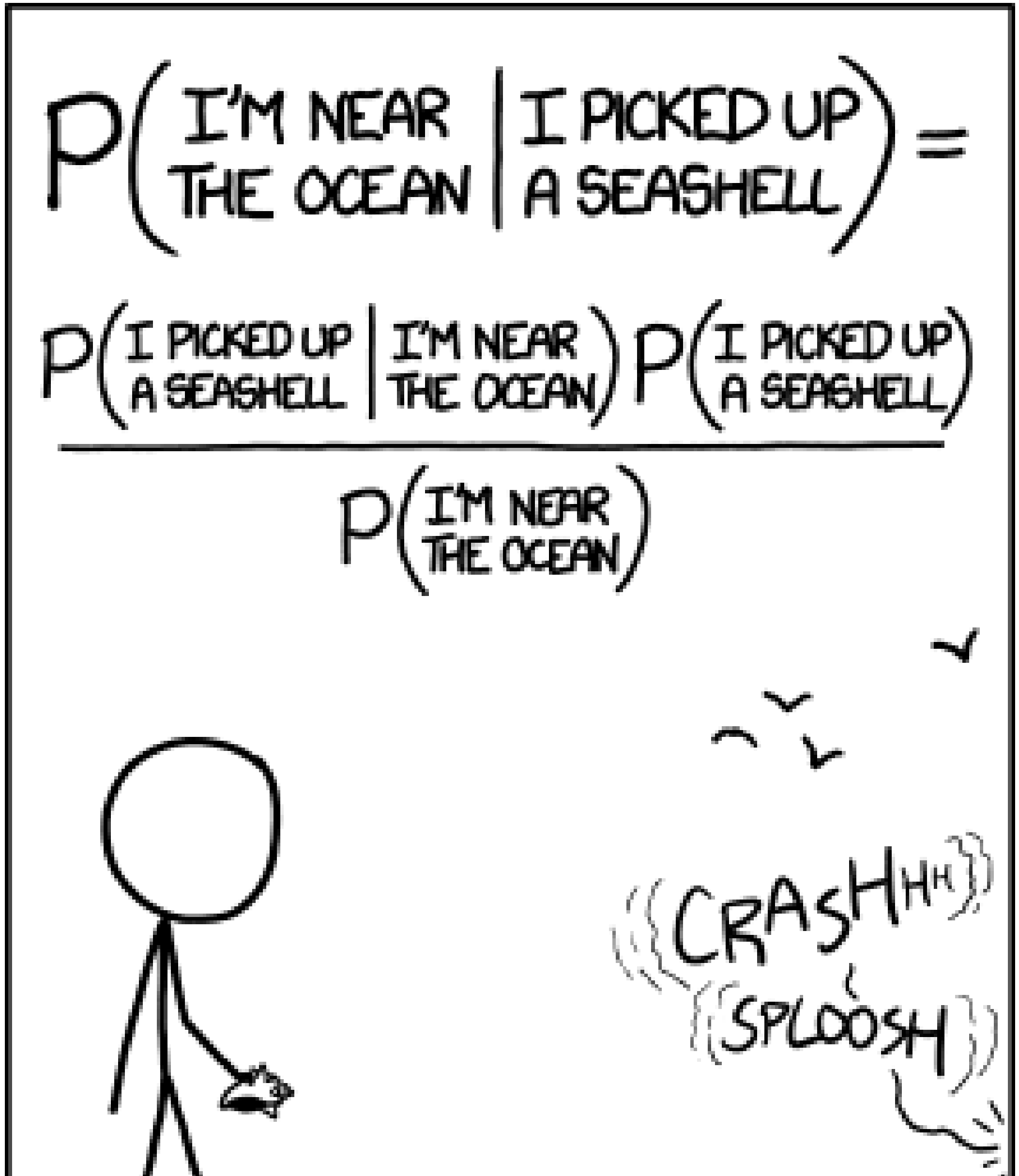
$P_1 \mid$ open 3 $= 1/3$

$P_2 \mid$ open 3 $= 2/3$

$P_3 \mid$ open 3 $= 0$

Monty Hall Problem





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

$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B)$$

$$\Pr(A|B) = \Pr(B|A)\cdot\Pr(A) / (\Sigma \Pr(B|A)\cdot\Pr(A))$$

Factoring probabilities:

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

$$= \Pr(A|B,C)P(B|C)\Pr(C)$$

BAYES RULE: alternate form

$$\Pr(A \mid B) = \Pr(B \mid A) \cdot \Pr(A) / \Pr(B)$$

$$\Pr(A|B) = \Pr(B|A)\cdot\Pr(A) / (\Sigma \Pr(B|A)\cdot\Pr(A))$$

Factoring probabilities:

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

$$= \Pr(A|B,C)P(B|C)\Pr(C)$$

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

Cumulative distribution function (cdf):

$F(z) = Pr(Z \leq z_k) = \Sigma$ $f(z)$ 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):

$F(z) = \Pr(Z \leq z)$ where $0 \leq F(z) \leq 1$

$\Pr(Z = z)$ is infinitely small

$$\begin{aligned}\Pr(z \leq Z \leq z + dz) &= \Pr(Z \leq z+dz) - \Pr(Z \leq z) \\ &= F(z+dz) - F(z)\end{aligned}$$

Probability density function (pdf):

$f(z) = dF/dz$

$f(z) \geq 0$ but NOT bound by 1

Continuous Random Variables

f is derivative of F
F is integral of f

$$Pr(z \le Z \le z+dz) = \int_{z}^{z+dz} f(z)$$

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 ≥ 0

What are the values of F(z) and f(z):
At z = 0 ?
As z $\rightarrow$ ∞ ?

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

Exponential

$$\mathrm{Exp}(x|\lambda) = \lambda \exp(-\lambda x)$$

At z = 0

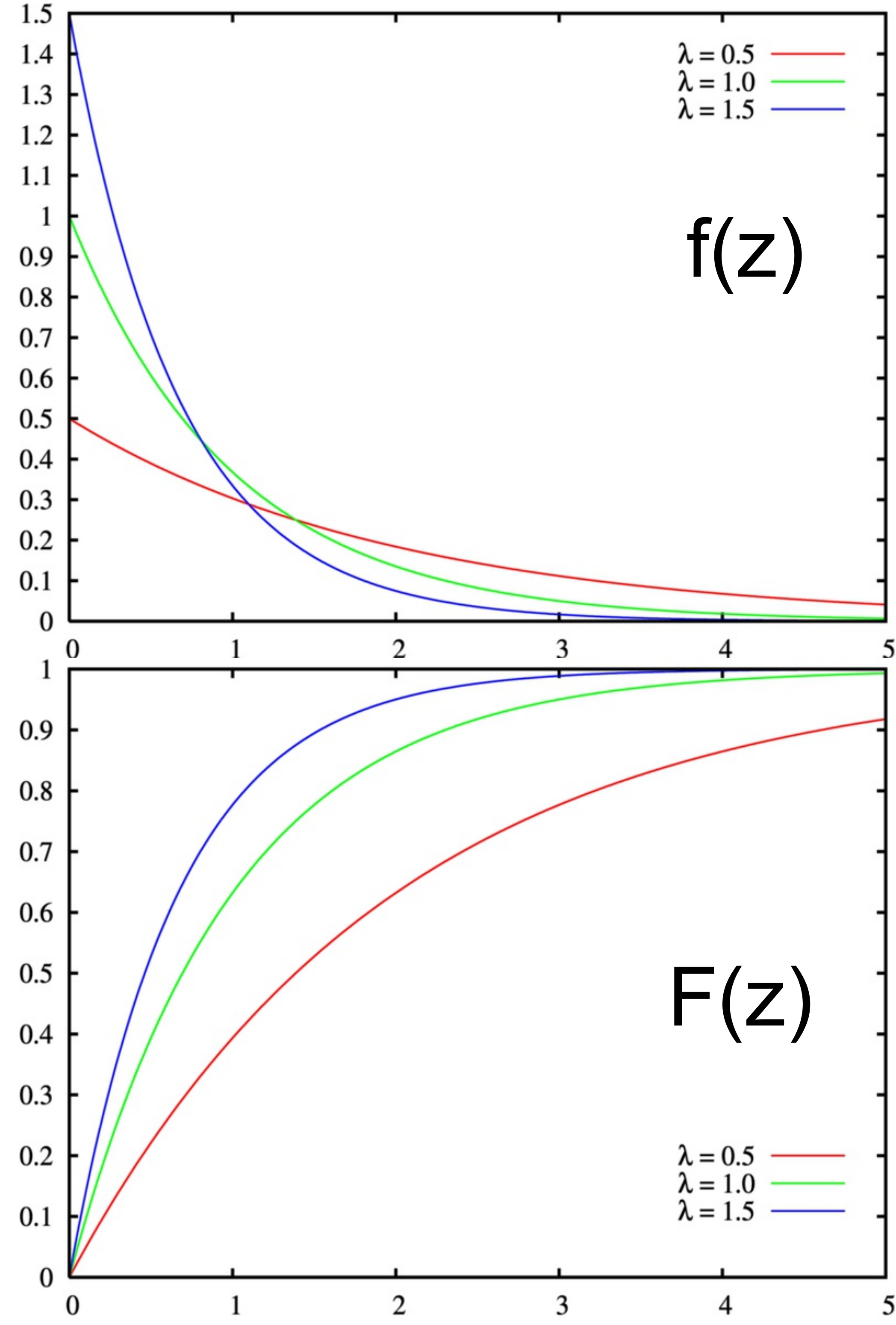
f(z) = Θ

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

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

$$= -x\exp(-\lambda x)\Big|_0^\infty + \frac{1}{\lambda}\int_0^\infty \lambda \exp(-\lambda x)$$

$$E[x] = 1/\lambda$$

Properties of means

$E[c] = c$

$E[x + c] = E[x] + c$

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

$E[x+y] = E[x] + E[y]$

(even if X is not independent of Y)

$E(xy) = E[x]E[Y]$ **only** if independent

$E[g(x)] != g(E[x])$ **Jensen's Inequality**

Central Moments

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

Second Central Moment = Variance = σ^2

Properties of variance

$$Var(aX) = a^2 Var(X)$$

$$Var(X+b) = Var(X)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2\mathrm{Cov}(X, Y)$$

$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2\mathrm{abCov}(X, Y)$$

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

$$Var(X) = Var(E[X|Y]) + E[Var(X|Y)]$$

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