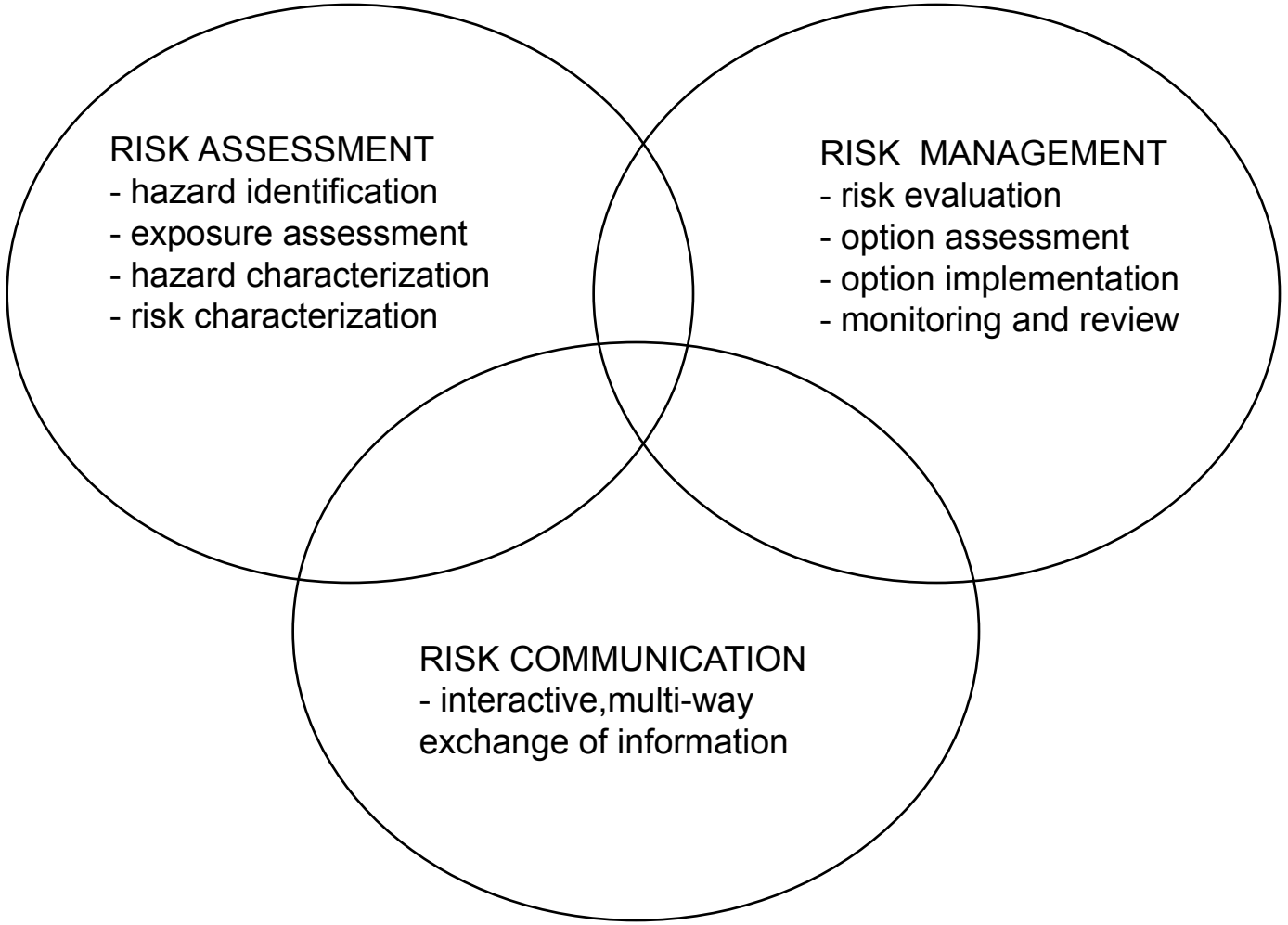
Quantitative Microbial Risk Assessment
SAFEGRO Project 2024

GREG PAOLI

RISK SCIENCES INTERNATIONAL
OTTAWA, CANADA

PRESENTED AT HANOI, OCT. 28-31, 2024

Risk Analysis:
The Org-Chart View of Risk Management



Codex Alimentarius Commission System

Quantitative Risk Assessment

DETERMINISTIC MODELING

What is a deterministic model?

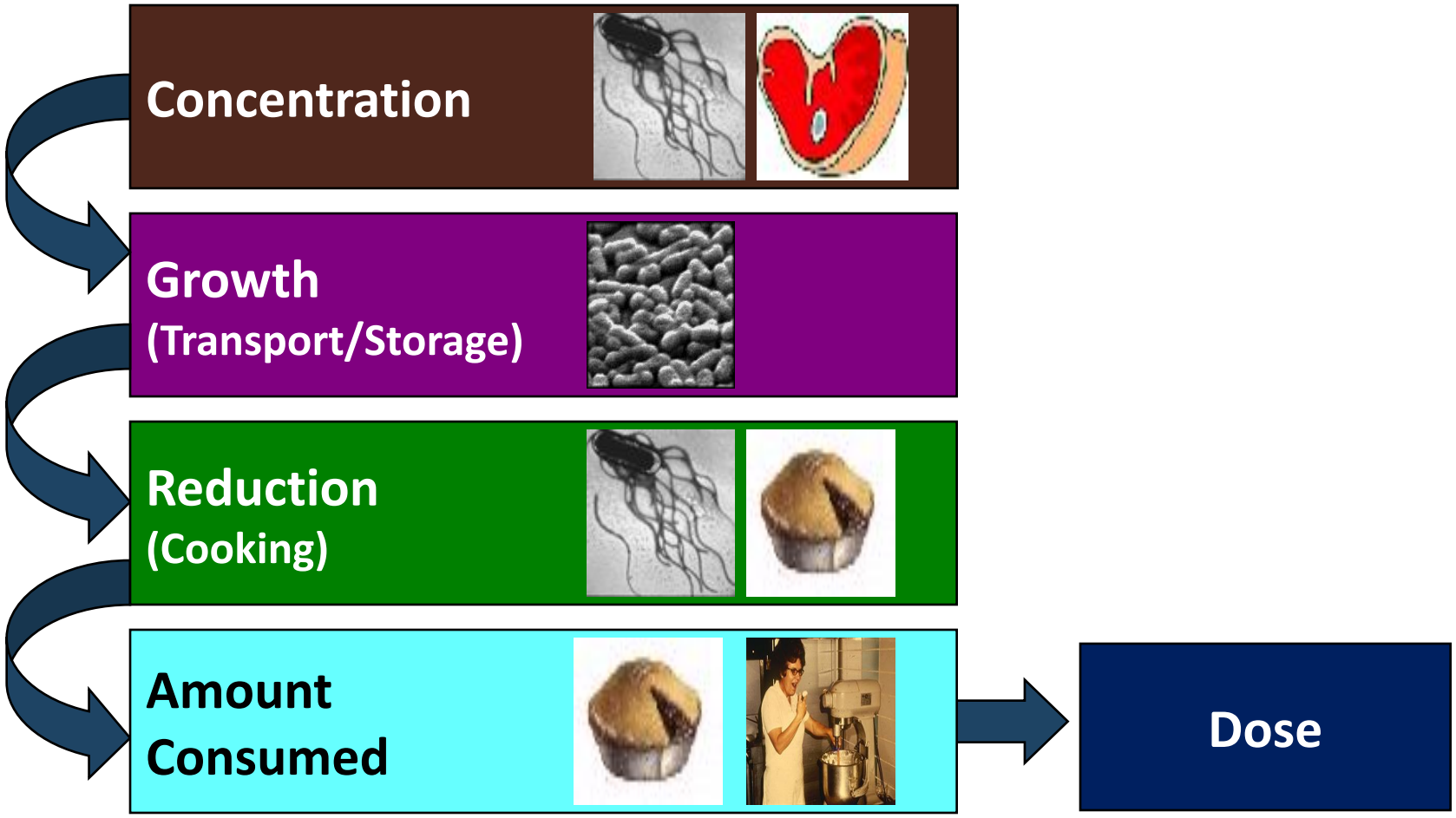
- In a deterministic model, the outcomes are precisely determined through known relationships among model parameters
- A given input will always produce the same output
- There is no consideration of any random variation in the system
- Model can be built using expected values, worst case estimates, etc.

QMRA Scenario: Building a deterministic model

- Bug “X” is present in Meat Pies
- Bug “X” can grow in Meat Pies
- Bug “X” can be inactivated by cooking

- Meat Pies are consumed by college students
- college students occasionally store the Meat Pies improperly
- college students sometimes do not cook Meat Pies well enough.

Example Scenario



Example Scenario

- First Approach
  - Estimate dose using mean values
- Second Approach
  - Estimate dose using worst case

Example Scenario

Mean Values

- Bug “X” Concentration = 2.0 log CFU/g
- Bug “X” Growth = 1.5 log (unitless multiplier)
- Bug “X” Inactivation = 3.6 log (unitless mulitiplier)
- Serving Size = 53.33 g

Example Scenario

Worst Case (upper limit)

- Bug “X” Concentration = 4.0 log CFU/g
- Bug “X” Growth = 1.85 (unitless multiplier)
- Bug “X” Inactivation = 2.6 log (unitless multiplier)
- Serving Size = 85.00 g

Point Estimate Results

- Mean Values
  - Estimated Dose Ingested
  - Approx. 36 organisms

Calculation

$(10^{[2 + 1.5 - 3.6]} \times 53.33)$

- Conservative Values
  - Estimated Dose Ingested
  - Approx. 152,000 organisms

Calculation

$(10^{[4 + 1.85 - 2.6]} \times 85.00)$

Discussion

- If illness is very unlikely with doses below 1,000 organisms, but increases above 1,000, are meat pies a “Safe Food”?
- Why could you argue they are NOT safe?
- Why could you argue they are safe?

Interpreting Point Estimates

- If conservative point estimate falls below maximum acceptable risk, then we know that the risk is truly acceptable
  - … but the extent of overprotection is unknown

- If conservative point estimate falls above maximum acceptable risk, then we do not know if the risk is truly unacceptable or is the result of propagated conservatism.

Burmaster 1995

Quantitative Risk Assessment

CASE STUDY:
CRONOBACTER SAKAZAKII
IN POWDERED INFANT FORMULA

Case study: Complex deterministic model

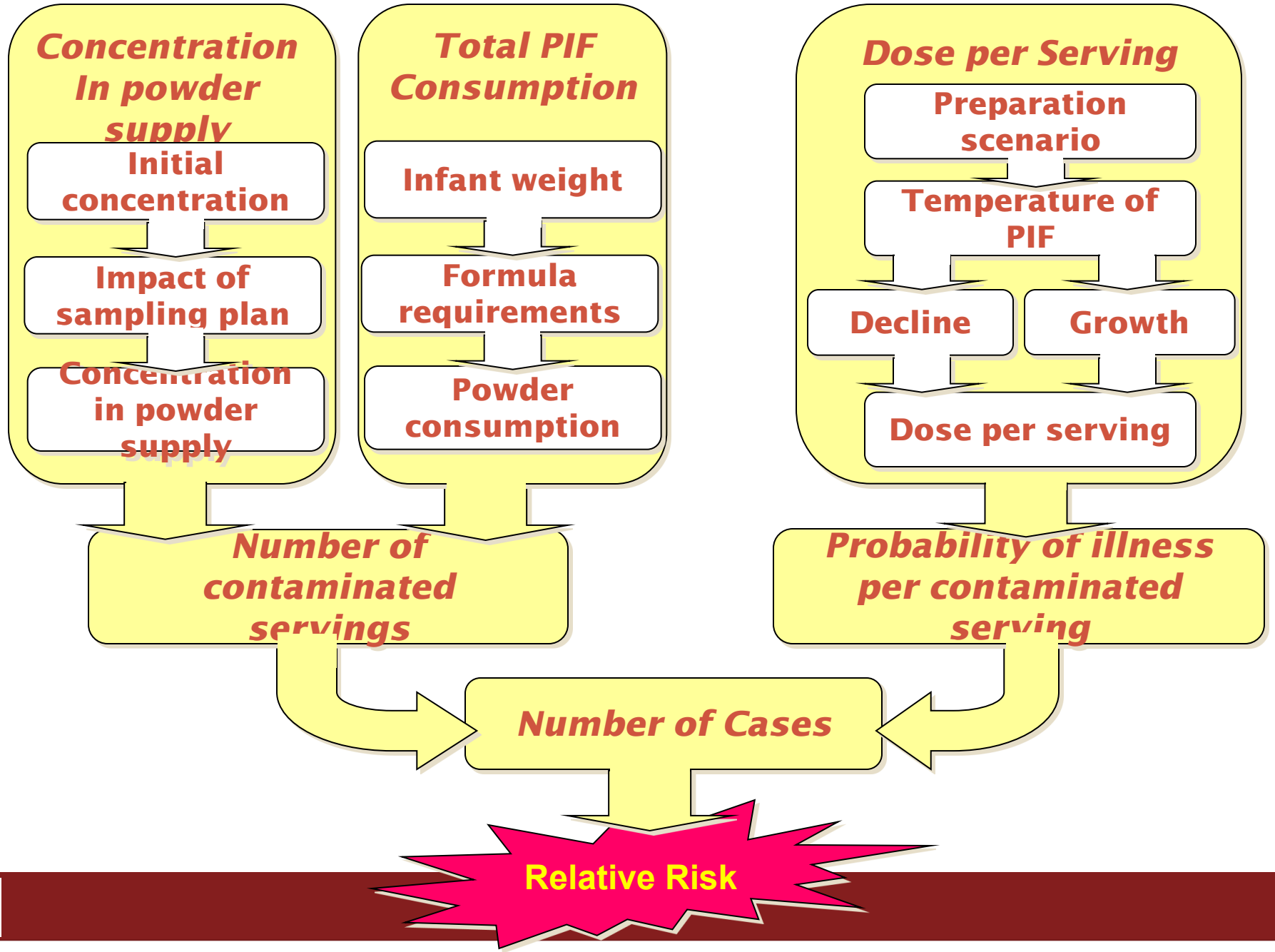
- *Cronobacter sakazakii* in powdered infant formula
- Based on an actual risk assessment tool that is publicly available
- Illustrates a real-world example of a deterministic model used in risk-based decision making
- Incorporates features common to the application of microbiological risk assessment in many domains
  - Dealing with predictive microbiology

C. Sakazakii in Powdered Infant Formula

- Powdered Infant Formula (PIF) that meets existing international/Codex standards has been implicated in cases of illness with *C. sakazakii*
- Codex therefore began the process of revising the code
  - Recommended International Code of Hygienic Practice for Food for Infants and Children
- At the request of FAO/WHO this risk assessment tool was completed
  - Provide risk-based scientific advice to Codex, and other risk managers, on the issue of *C. sakazakii* in PIF
  - Intended for 'live' use by risk managers in consultation with scientific working groups

Brief Summary of the Risk Assessment

- Model estimates the dose of *C. sakazakii* in prepared PIF at consumption, and subsequently the risk
- Specification of scenarios underpins the prediction of the dose in prepared formula at consumption
- Outputs are in terms of the change in relative risk across scenarios


Concentration In powder supply
Initial concentration
Impact of sampling plan
Concentration in powder supply
Total PIF Consumption
Infant weight
Formula requirements
Powder consumption
Dose per Serving
Preparation scenario
Temperature of PIF
Decline
Growth
Dose per serving
Number of contaminated servings
Probability of illness per contaminated serving
Number of Cases
Relative Risk

Describing Preparation Scenarios

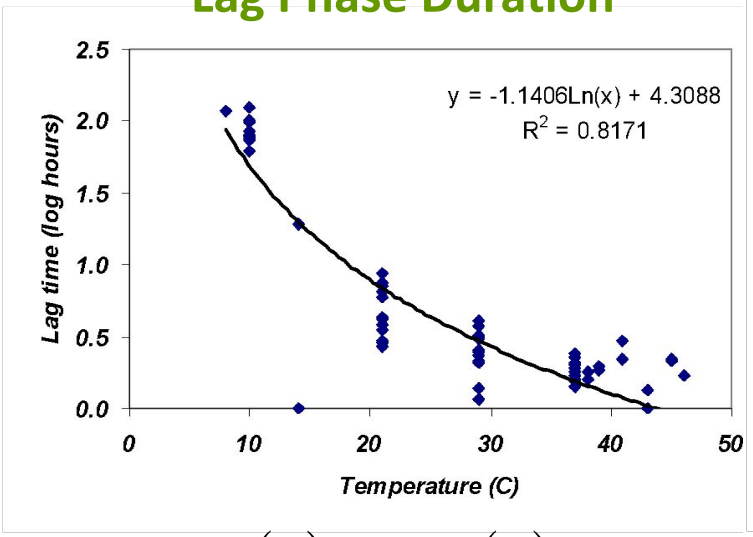
- Scenarios are defined in terms of preparation, cooling/holding, re-warming and feeding
- Scenarios consider:
  - Temperature of re-hydration liquid
  - Preparation scenario (single bottle, 1litre container..)
  - Temperature for cooling/holding
  - Room temperature for feeding
  - Duration of each preparation stage
- Model predicts the temperature of the formula over entire time from re-hydration to feeding

A Closer Look at Preparation and Handling

Preparation Scenario

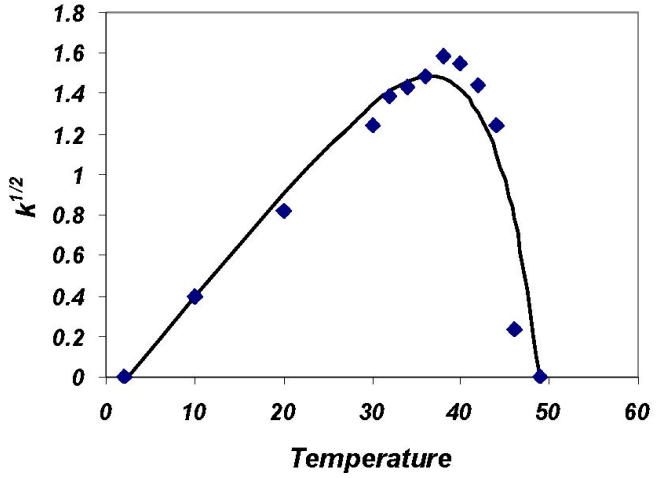
Temperature of PIF

Lag Phase Duration



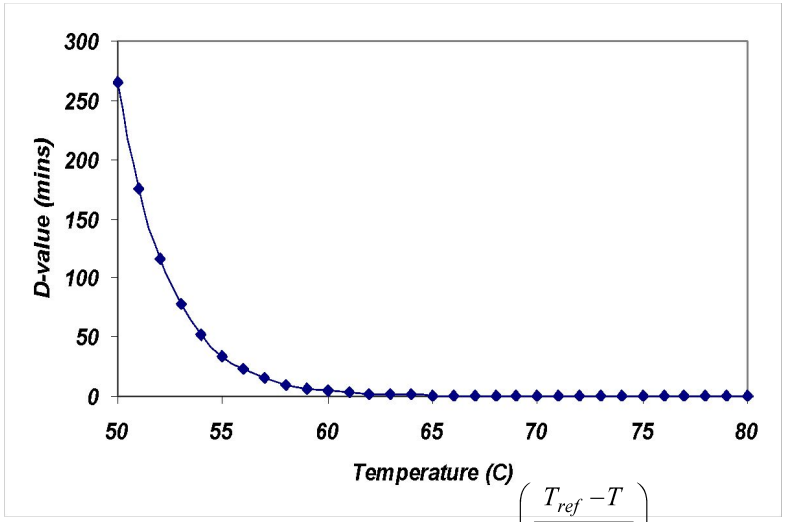
$$Log_{10}(\lambda) = c_L \ln(T) + b_L$$

Growth Predictions



$$\sqrt{k} = b_G(T - T_{\min})\{1 - \exp(c_G(T - T_{\max}))\}$$

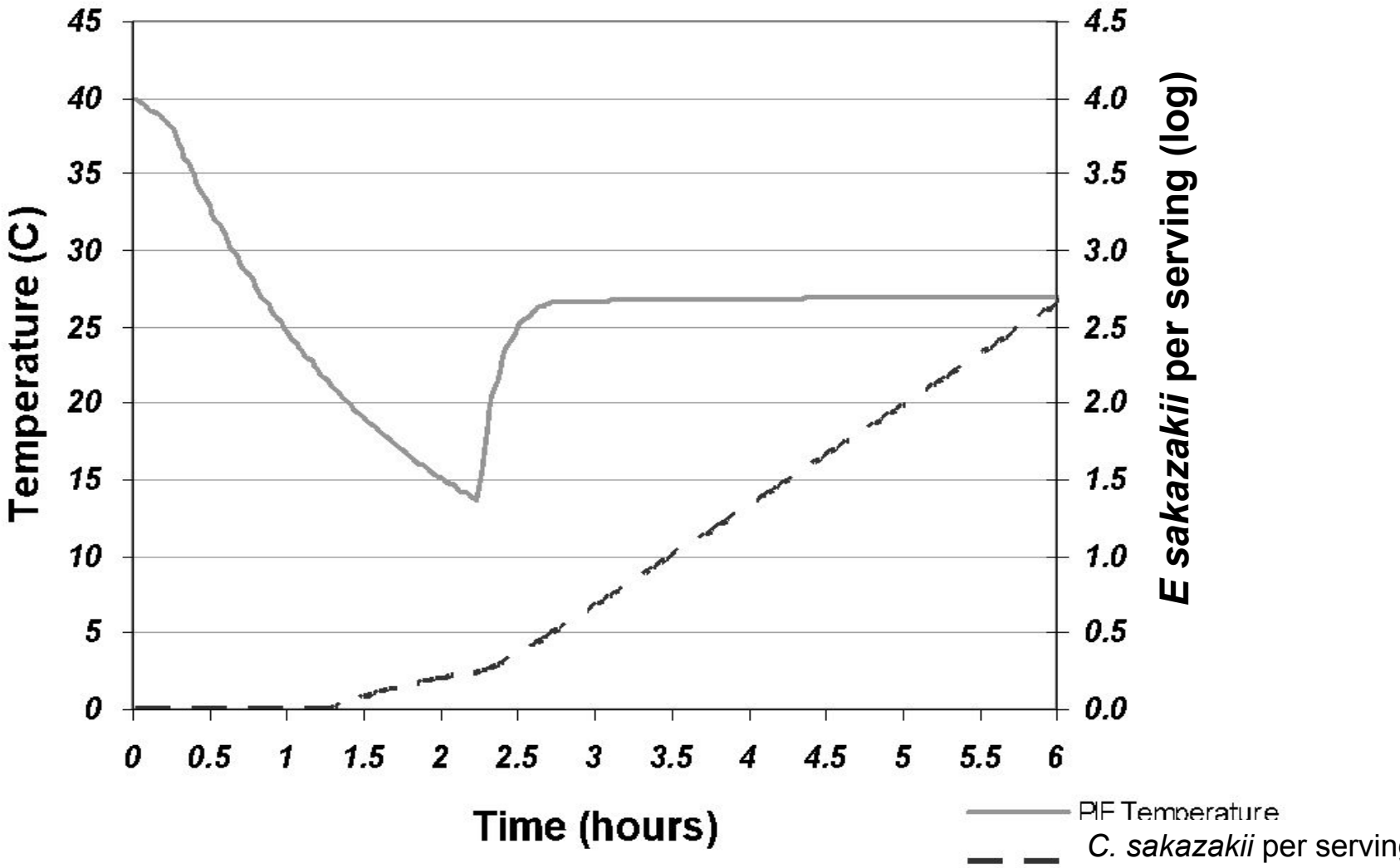
Decline Predictions



$$D_T = D_{ref} - 10^{\left(\frac{T_{ref} - T}{Z - value}\right)}$$

Dose per serving

Population change over time



Comparing Preparation & Handling Scenarios

- FAO/WHO convened an expert meeting
  - January 2006, Rome
- Scenarios were created by a working group at the meeting
- Questionnaires were sent to hospitals around the world
- An extensive list of scenarios was explored
  - e.g. refrigerator temperature/time, room temperature…
- Results were generated at the meeting and interpreted by the working groups
  - Full report available on JEMRA website
    - http://www.who.int/foodsafety/publications/micro/mra10.pdf

Basic Scenarios

- Eight basic scenarios were investigated
- Conditions were specified for cool, warm and very warm room temperatures
- Scenarios covered the combinations of:
  - Cooling by refrigeration (4°C) or holding at room temperature
  - Inclusion or exclusion of an explicit re-warming action
  - Short or long feeding periods
- Each of these scenarios was run at a series of different reconstitution temperatures
  - 10, 20, 30, 40, 50, 60 and 70°C
  - Resulting in the comparison of 168 different preparation scenarios

Example Output: Basic Scenarios

Table. 11. Relative risk of different preparation, storage and handling practices for formula prepared and used at a warm ambient room temperature (Room temperature = 30°C) (+ X means an increase in risk of X fold, - X means a decrease in risk of X fold)

Preparation, storage and feeding scenarios	Relative increase or decrease in risk compared to the baseline scenario of 1 at different temperatures of rehydration of PIF						
	10 °C	20 °C	30 °C	40 °C	50 °C	60 °C	70 °C
Refrigeration, re-warming, extended feeding period	**+ 2**	**+ 34**	**+ 8**	**+ 27**	**+ 83**	**+ 1.8**	> - 100,000
Refrigeration, re-warming, short feeding period	1	1	1	1	**+ 2.6**	- 1.3	> - 100,000
Refrigerated storage, no re-warming, extended feeding period	1	1	1	1	**+ 2.7**	- 1.3	> - 100,000
Refrigeration, no re-warming, short feeding period	1	1	1	1	1	- 1.3	> - 100,000
No refrigeration, re-warming, extended feeding period	**+ 3**	**+ 6**	**+ 15**	**+ 55**	**+ 161**	1	> - 100,000
No refrigeration, re-warming, short feeding period	1	1	1	**+ 1.7**	**+ 5**	- 1.3	> - 100,000
No refrigeration, no re-warming, extended feeding period	1	1	**+ 2.8**	**+ 22**	**+ 97**	- 1.3	> - 100,000
No refrigeration, no re-warming, short feeding period	1	1	1 (Base line)	1	**+ 3**	- 1.3	> - 100,000

Example Output: Refrigeration

Table 18. Comparison of the relative risk related to holding time in /out of refrigeration before extended feeding (2 hour) for scenarios conducted at a warm ambient room temperature. (+ X means an increase in risk of X fold).

	Time between preparation and feeding	Relative increase in risk compared to the baseline scenario of 1 at different temperatures of rehydration of PIF						
		10°C	20°C	30°C	40°C	50°C	60°C	70°C
Refrigeration	2 hour (baseline)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)
	4	**+1.06**	**+1.09**	**+1.09**	**+1.12**	**+1.17**	**+1.19**	1
	6	**+1.08**	**+1.11**	**+1.11**	**+1.14**	**+1.19**	**+1.22**	1
	8	**+1.11**	**+1.11**	**+1.13**	**+1.16**	**+1.19**	**+1.24**	1
No Refrigeration	2 hour (baseline)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)	1 (Base line)
	4	**+11**	**+16**	**+32**	**+46**	**+53**	**+63**	1
	6	**+354**	**+600**	**+1,189**	**+1,377**	**+680**	**+2,745**	1
	8	**+ 12,721**	**+ 18,548**	**+ 15,268**	**+2,793**	**+705**	**+ 65,502**	

Providing Advice

- Following use of the risk assessment the meeting concluded that:
  - Some of the current instructions on PIF product labels, and those recommended by health authorities, may lead to increased risk of *C. sakazakii* illnesses, and that these should be reviewed in light of the risk assessment results
- The assessment has been used by FAO/WHO to develop guidance, and these are publicly available
  - Guidelines for the safe preparation, storage and handling of powdered infant formula
  - http://www.who.int/foodsafety/publications/micro/pif2007/en/
  - Tool available freely online at www.fstools.org

Quantitative Microbial Risk Assessment

HAZARD AND RISK CHARACTERIZATION: MICROBIAL DOSE-RESPONSE MODELS AND ESTIMATING THE NUMBER OF ILLNESSES

Hazard Characterization in MRA

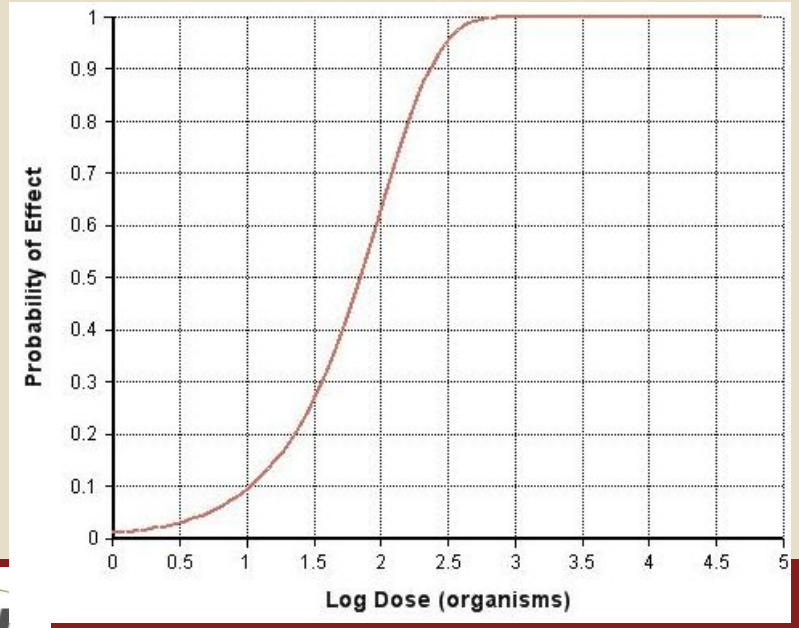
- The qualitative(?) and/or quantitative evaluation of the nature of the adverse health effects associated with the biological agent

- Should explicitly consider the complexity of the interaction (including sequelae) between human and agent following exposure as well as the potential for further spread
  - Dose-response assessment should be performed

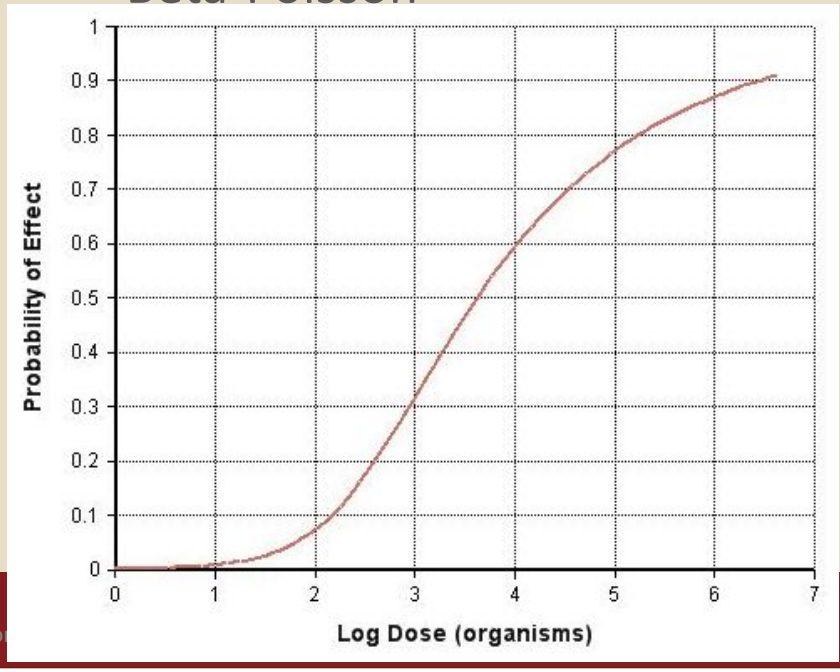
Dose-Response Assessment

- Dose response models are mathematical functions that describe the dose response relationship for specific pathogens, transmission routes, and hosts
- Estimate the risk of a response (for example, infection, illness or death) given a known dose of a pathogen

•Exponential



Beta-Poisson

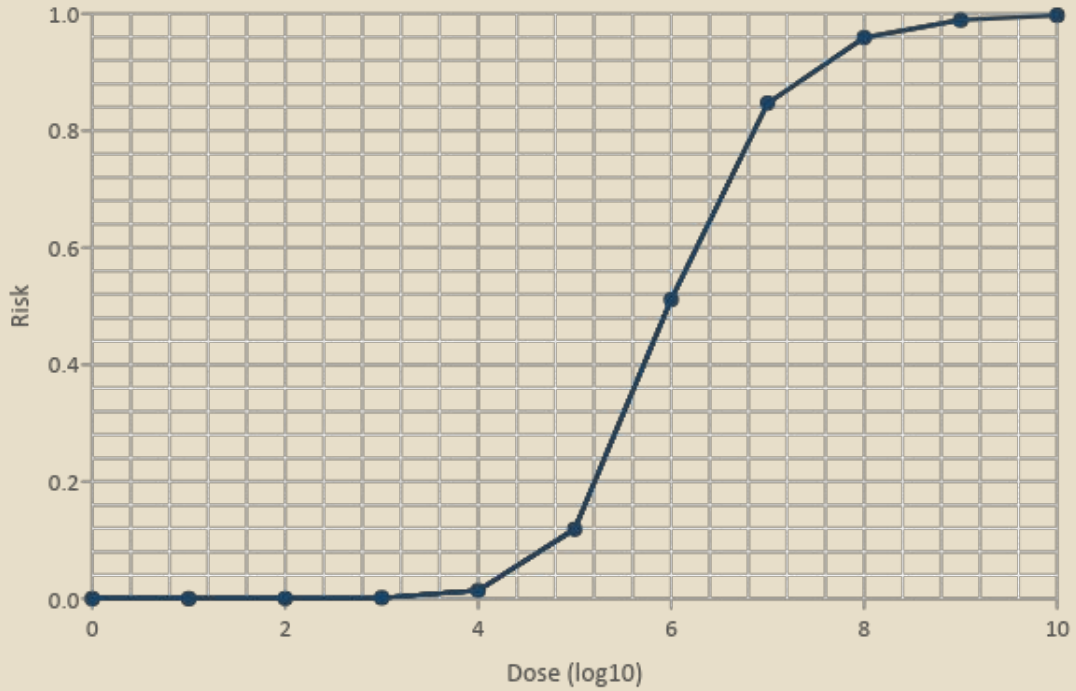


Microbial Dose Response Models

- Always considered to be acute exposure
- 1 CFU is capable of causing infection
  - Theory of minimum infectious dose (MID) no longer accepted
- May be based on feeding studies or outbreak data

Completing the Meat Pie Model

- Assume Bug X follows a Beta-Poisson dose-response relationship and add it to the deterministic model
  - P = $P_{ill} = 1 - \left(\frac{1+d}{\beta}\right)^{-\alpha}$
  - α = 0.581, β= 4.11 x 10^5



Key Resource for Microbial Dose-Response Models

- Large compendium of experiments and models compiled:
  - https://qmrawiki.org/framework/dose-response/experiments

- See also, https://www.who.int/publications/i/item/9789240024892

TABLE 7. Dose–response models and parameter estimates commonly used in QMRA

Organism	Reference	Model	Parameters	Lower bound (Percentile)	Upper bound (Percentile)
Salmonella spp.	FAO/WHO (2002a)	Beta-Poisson	α=0.1324 β=51.43	0.0940 (2.5th) 43.75 (2.5th)	0.1817 (97.5th) 56.39 (97.5th)
Listeria monocytogenes [a]	FAO/WHO (2004)	Exponential (susceptible) Exponential (healthy)	$r=1.06\times10^{-12}$ $r=2.37\times10^{-14}$	2.47×10^{-13} (5th) 3.55×10^{-15} (5th)	9.32×10^{-12} (95th) 2.70×10^{-13} (95th)
Campylobacter spp.[b]	FAO/WHO (2009d)	Beta-Poisson	α=0.21 β=59.95		
Shigella dysenteriae/ *E. coli* O157	Cassin *et al.* (1998)	Beta-binomial	α=0.267 β=Lognormal (5.435, 2.47²)		
Vibrio vulnificus	FAO/WHO (2005)		$\alpha=9.3\times10^{-6}$ β=110 000		

[a] For *L. monocytogenes*, newer animal model data (Roulo *et al.*, 2014; Smith *et al.*, 2003, 2008; Williams *et al.*, 2007, 2009) and outbreak data (Pouillot *et al.*, 2016) suggest much higher *r*-values and hence lower ID_{50} values than predicted by this model which was based on the method of Buchanan *et al.* (1997) of matching expected loads of *L. monocytogenes* across the food supply to the total annual cases in a community, and which relies on many untested assumptions.
[b] The dose–response relation is for infection. The conditional probability of disease following infection was 33 percent (29/89) and can be described by a beta(30,61) distribution.

Risk Characterization: Getting to Number of Illnesses

- Basic equation
  - Probability of Illness: $P_{ill} = G(d)$
  - G(d) is a function that converts a dose into a probability of illness given a single serving containing the dose, d.
- When doses are variable:
  - Probability of Illness: $MeanP_{ill} = \int_{d_{min}}^{d_{max}} G(d) f(d) dd$
  - Number of illnesses: $Nill = N_{servings} \times MeanP_{ill}$
  - G(d) is a function that converts a dose into a probability of illness given that dose.
  - $f(d)$ is a probability distribution of the variability of doses in servings of food
  - The distribution *f(d)* and the integral is typically generated using Monte Carlo Simulation

Quantitative Risk Assessment

REVIEW OF PROBABILITY AND INTRODUCTION TO PROBABILISTIC MODELING

Definition of Probability

- Random experiment may reveal a pattern
  - Pattern of heads and tails
  - Pattern of different heights of people
- Repeat experiment large number of times to learn more about pattern
- Relative frequency of particular outcome compared to other outcomes tends toward constant value

Probability

Example

- Probability is a measure between 0 and 1
- A ball is selected at random from a bag containing 3 red balls and 7 white balls
- The probability that a red ball will be drawn is 3/10

- At Random – means each of the 10 balls has same probability (chance) of being selected
  - All 10 outcomes are equally likely
- The probability that a white ball is drawn is 7/10
  - Total of two probabilities is 1 – no other outcome is possible – ball is either white or red

Sample Space

- Values that the outcome of random experiments can take
  - Heads or tails when tossing a coin
  - All possible heights of people in a room

- Subset of values
  - Heights between 150 cm and 155 cm

Event

Random experiment

- Process which yields information
  - result of tossing a coin
  - height of next person to enter room
  - time each slide will take to present
- If random then unsure of outcome
  - Will you get a head or a tail?
  - Unsure of height before seeing (and measuring) person
  - Some slides will be faster than others

Conditional probability

- A and B are outcomes of a random experiment

  - P(A) - Probability A occurs
  - P(not A) - Probability A does not occur
  - P(A ∩ B) - Probability A *and* B occur
  - P(A ∪ B) - Probability either A or B or both occur
  - P(A | B) - Probability A occurs given B has occurred

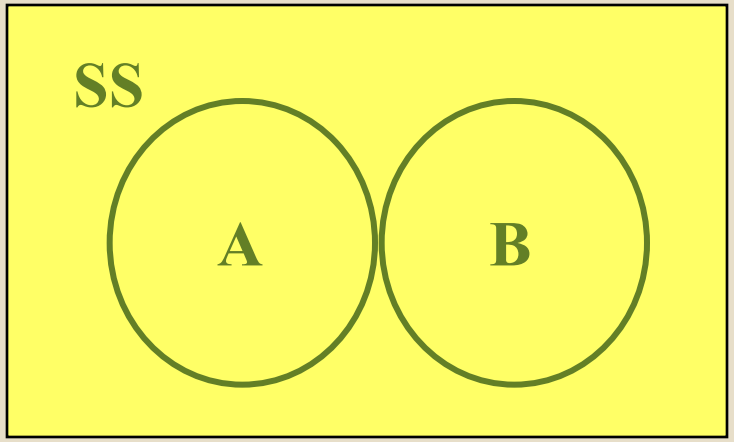
Representations of Probability

- Help consider a problem visually
- Often prevent simple mistakes

  - Venn Diagrams
  - Event Trees

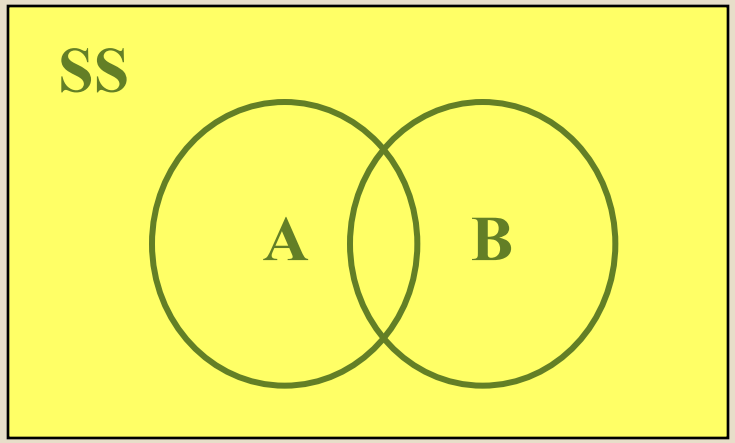
Venn Diagrams

- Venn diagrams show how the sample space is divided into events
- Square is total sample space = 1
  - P(not A) = 1 - P(A)
  - P(not B) = 1 - P(B)

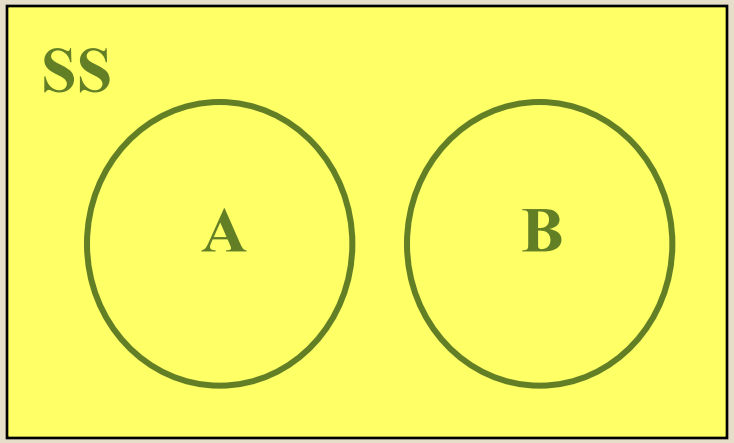


Venn diagrams

- mutually exclusive events cannot occur at the same time

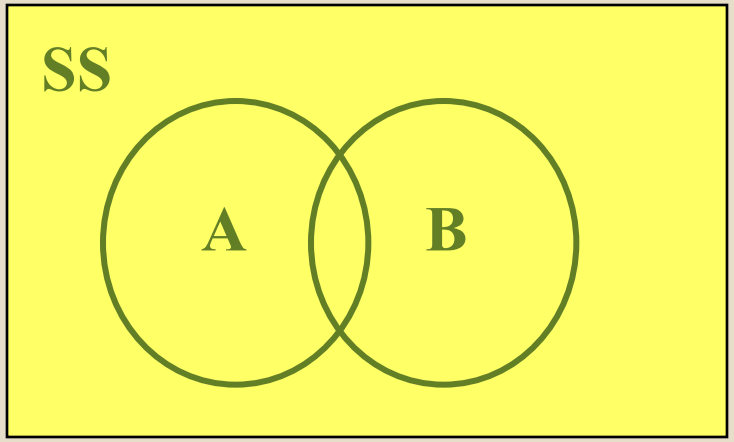


- A and B not mutually exclusive



- A and B mutually exclusive

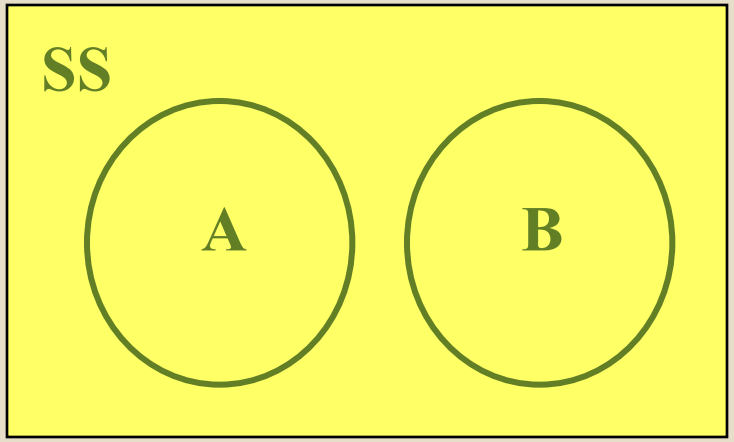
Calculations



If A and B are **not** mutually exclusive

- $P(A \cap B) = P(B|A) \times P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Calculations



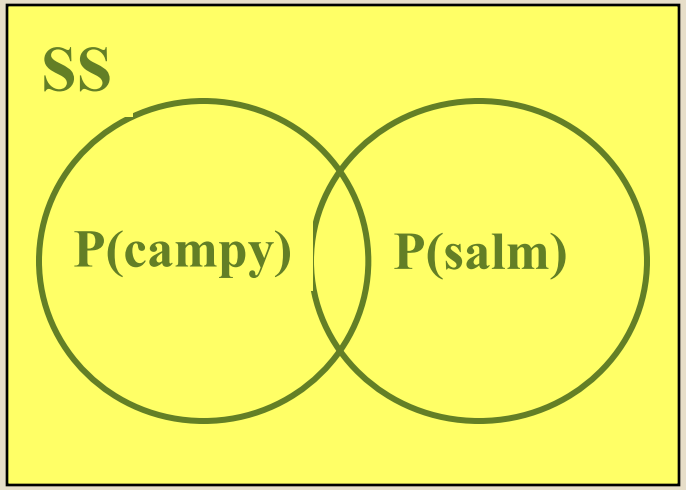
If A and B are mutually exclusive

- $P(A \cap B) = P(B|A) \times P(A) = 0$ (why zero?)
- $P(A \cup B) = P(A) + P(B)$

Venn Diagram example

- A chicken pie has 20% chance of having campylobacter, and 10% chance of having salmonella. What is the probability that a chicken pie has either campylobacter or salmonella, or both?
- P(campy) = 0.2
- P(salm) = 0.1
- P(campy & salm) = 0.2+0.1-(0.2*0.1)=0.28

Example

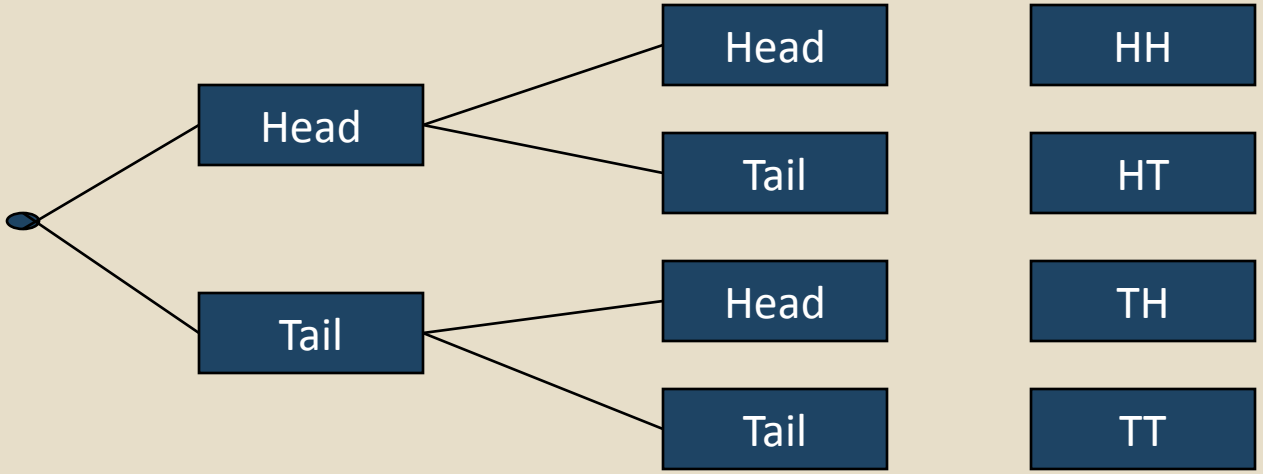


P(campy) = 0.2

P(salm) = 0.1

P(campy $\cup$ salm) = 0.2+0.1-(0.2*0.1) =0.28

Event Tree (two coin flips)



- P(0 Heads) = (0.5 × 0.5) = 0.25
- P(1 Head) = (0.5 × 0.5) + (0.5 × 0.5) = 0.5
- P(2 Heads) = (0.5 × 0.5) = 0.25

p(x=0,1,2) = {0.25, 0.5, 0.25}

PROBABILITY DISTRIBUTIONS

What is a Probability Distribution

- A function that describes all the values that a random variable can take, and the probability associated with each
- Random variable – must take one and only one value from sample space at any time
- Values in sample space are mutually exclusive
- Probabilities in distribution sum to 1

Probability distributions can be …

- Discrete
- Continuous
- Parametric
- Non-parametric

Discrete distributions

- A function that can take a discrete number of values (not necessarily finite).
- Each value (x) has exact probability of occurrence
- Sum of probabilities equals unity

$$\sum_j P_j = 1$$

- This is most often the non-negative integers
- Often referred to as a probability mass function

Discrete example

- Examples of discrete variables are
  - Outcome of tossing a coin (either H or T)
  - Gender (M or F)
  - Number of cars in a parking lot (integer)
  - A sample result that is either positive or negative
  - Number of organisms (0,1, 2, …)
  - Health status (immunocompromised, normal)
  - Day of the week that an event occurs

Continuous distributions

- Are defined for an infinite number of points over a continuous interval
- Area under curve equals unity
- Probability for any particular value is zero
- The probability that x is between two points a and b is

$$p[a \le x \le b] = \int_a^b f(x)\,dx$$

- Often referred to a probability density function

Continuous examples

- Examples of continuous variables are
  - Time, Duration, and Intensity of Rainfall
  - Length, Weight, Height
  - Position of an accident (pipeline rupture)
  - Temperatures
  - Concentrations, Volumes, Rates
  - Distance an ambulance must travel to rescue

Borderline Cases

- Some quantities inherently *discrete*, but characterized as *continuous* for computational convenience:
  - Large numbers of pathogens
  - The size of an exposed population
  - Number of phone calls handled by an emergency dispatcher in a week
- Impact of this choice can range from trivial (often) to serious (rare, but important). Check it out!

Borderline Cases

- Some quantities are continuous, but may be characterized as discrete:
  - Reported measurements (rounded off)
    - Temperature?
    - Your height? Your age?
  - Building height (stories)
- If required, we can use various methods to recreate a continuous distribution from discrete data
  - Take into account the expected nature of the continuous phenomenon, the form of discretization (e.g. rounding), expected biases

They can also be…..

- Bounded
  - confined by 2 limits
- Unbounded
  - extends from ±infinity
- Partially bounded
  - constrained at one end

GRAPHICAL REPRESENTATIONS OF PROBABILITY

Example: discrete random variable

Example: continuous random variable

The cumulative distribution (cdf)

- The cdf is the probability that the variable takes a value less than or equal to x :

  For a continuous distribution

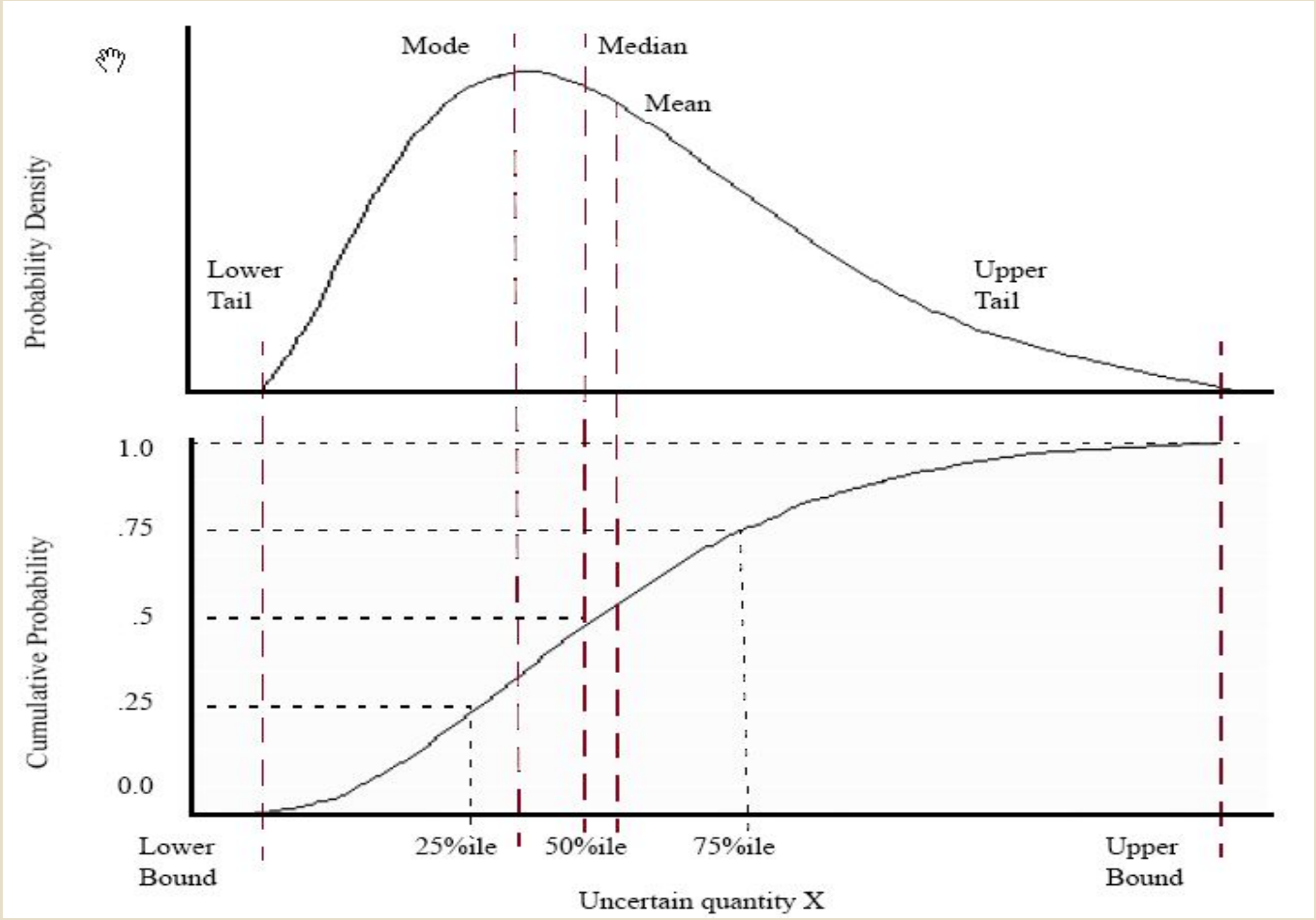
$$F(x) = \Pr_x[X \leq x]$$

$$F(x) = \int_{-\infty} f(\mu)\, d\mu$$

- For a discrete distribution

$$F(x) = \sum_{i=-\infty}^{x} f(i)$$

PDFs and CDFs



Review of statistical measures

- Measures of central tendency
  - Mean
  - Median
  - Mode
- Measures of dispersion
  - Range
  - Variance, Standard Deviation
- Percentiles

Measures of central tendency: Mean or average (m)

Discrete

$$m = \sum_{i=1}^{n} x_i p(x_i)$$

- n = number of possible outcomes in SS
- x_i = value of outcome *i*
- $p(x_i)$ = probability of outcome *i* occurring

Continuous

$$m = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- *x* = value of outcome
- *f(x)* = probability density function

Measures of Central Tendency

- **Mean**
- Toss a coin twice, how many heads?

- discrete distribution
- sample space SS={0,1,2}
- probability distribution p(x)={0.25,0.5,0.25}
- Mean = $0\times0.25 + 1\times0.5 + 2\times0.25 = 1$

Measures of Central Tendency

- Median
  - value that 50% of distribution is above and 50% of distribution is below
- Mode
  - most frequent observation or value with highest probability of occurrence (most likely value)

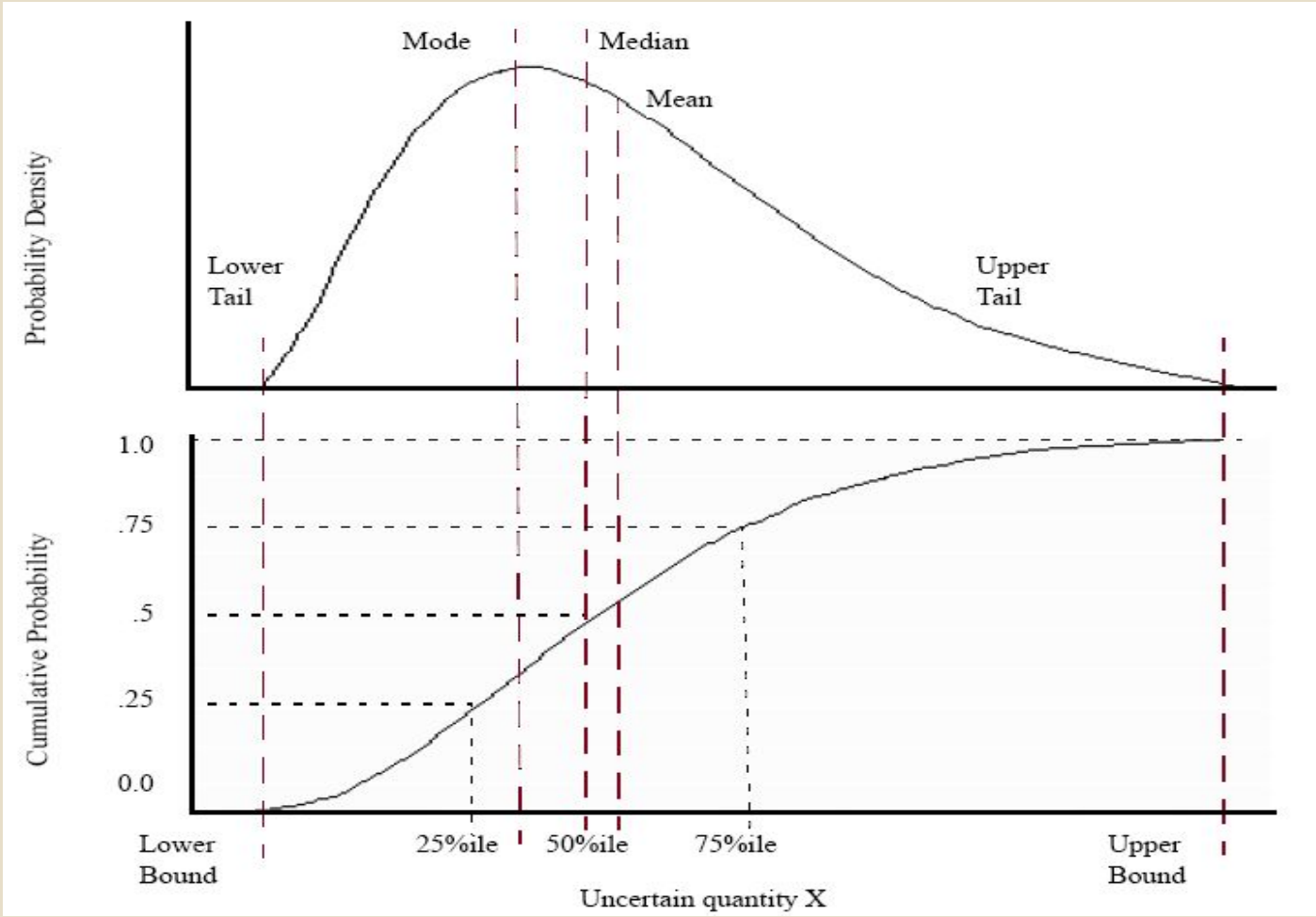
Measures of dispersion

- Range
  - difference between minimum and maximum values
  - example: range of deer calf weight is 69.3-25.8 = 43.5kg
- Variance
  - The average of squared deviations from the mean

Measures of dispersion

- Percentiles
  - xth percentile is value for which x% of the data has a lower value
  - also thought of in terms of “certainty”
    - 95% certain that number of sheep in a flock is less than 300
    - intervals of uncertainty
  - 50th percentile also called “median”

Measures of dispersion



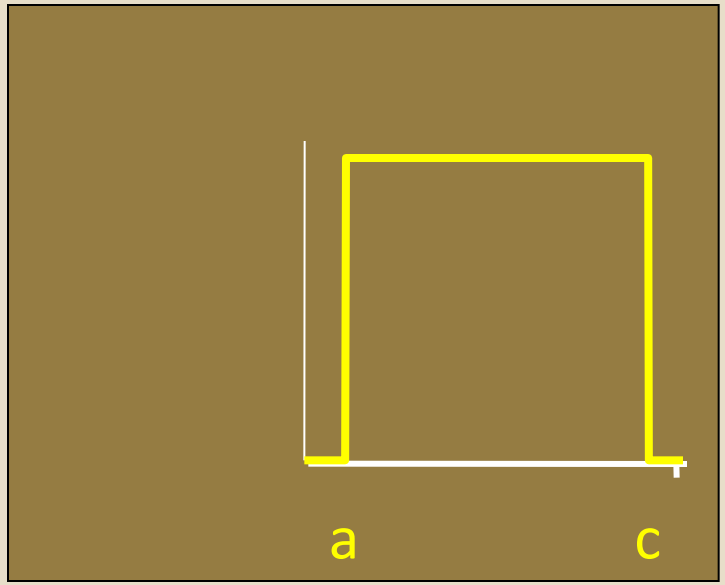
Quantitative Risk Assessment

A SIMPLE PROBABILITY DISTRIBUTION:

UNIFORM

Uniform(a,c)

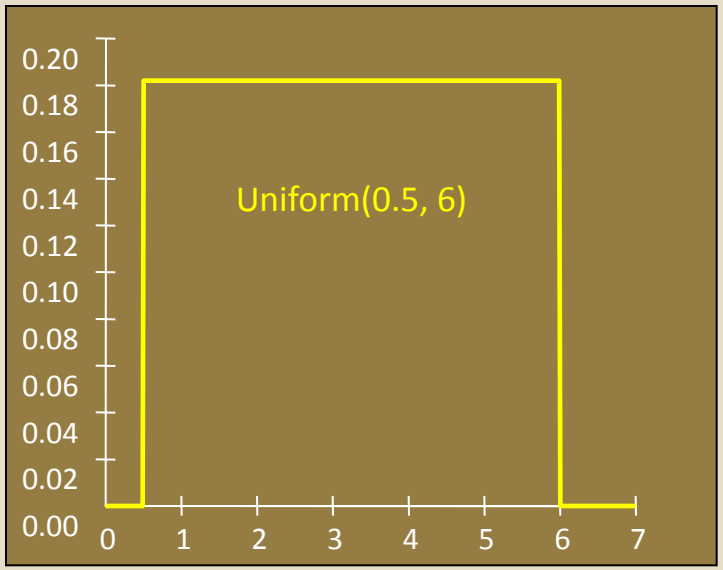
- Parameters: minimum (a); maximum (c)
- Assumes *all* values between a to c are equally likely to occur
- Often used to represent total ignorance
- continuous
- Bounded
  - Domain: ($a \leq x \leq c$)
- **Mean = (a+c)/2**



Example of Uniform Distribution

- The waiting time for treatment in the ER is not known, but can be between a minimum of 0.5 and a maximum 6 hours
- Distribution = Uniform(0.5,6)
- Mean = 3.25

=RiskUniform(a,c)



Quantitative Risk Assessment

INTRODUCTION TO MONTE CARLO SIMULATION

Monte Carlo simulation

- ***Simulation*** - any analytical method meant to imitate a real-life system, especially when other analyses are too complex mathematically or are too difficult or expensive to reproduce.
- For each probabilistic variable define the possible values with a distribution.
- Monte Carlo Analysis
  - systematically constructs the probability distribution of output variables, by randomly selecting values for input variables *according to their probability distributions*.

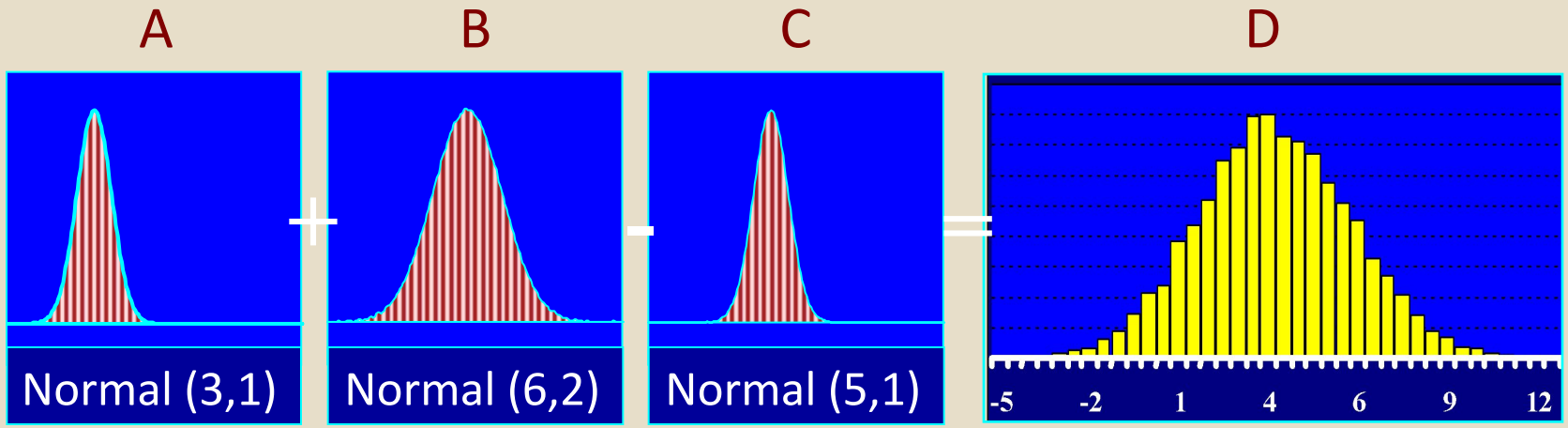
Monte Carlo Simulation

- The random selection process is repeated many times
  - multiple scenarios
- Each value represents one possible scenario
- Together, these scenarios give a range of possible solutions
- Some solutions are more probable and some less probable – probability distribution

Monte Carlo Simulation

- Monte Carlo analysis allows us to simulate variability and uncertainty in the values

Example : D= A+B-C



- Range of values for “D” and probability of occurring can be determined.

Monte Carlo Simulation

- When repeated many times the average solution will give an approximate answer to the problem
- Accuracy of this answer can be improved by simulating more scenarios.
  - More on this later in the course

Simulation Software: @Risk

- Demonstrate:
- How @Risk works with Excel
- Building a simple simulation model using @Risk
- Exploring probability distributions using @Risk
- Understanding @Risk output

Dice: A Stochastic Process

- Playing ‘Craps’: a simple example
- Determining the probability distribution for the sum (S) of two dice, X and Y
- Analytical Approach:

$$\mathrm{P}_{\mathrm{r}}[S=s]=\sum_{x}\mathrm{P}_{\mathrm{r}}[X=x]\mathrm{P}_{\mathrm{r}}[Y=s-x]$$

Dice: A Stochastic Process

- Roll the dice and take notes

Iteration	First Die		Second Die		Total
1	2	+	6	=	8
2	4	+	5	=	9
3	2	+	2	=	4
4	4	+	3	=	7
5	4	+	6	=	10

Winning: 7, 11 or Doubles

	1	2	3	4	5	6
1	D					7
2		D			7	
3			D	7		
4			7	D		
5		7			D	11
6	7				11	D

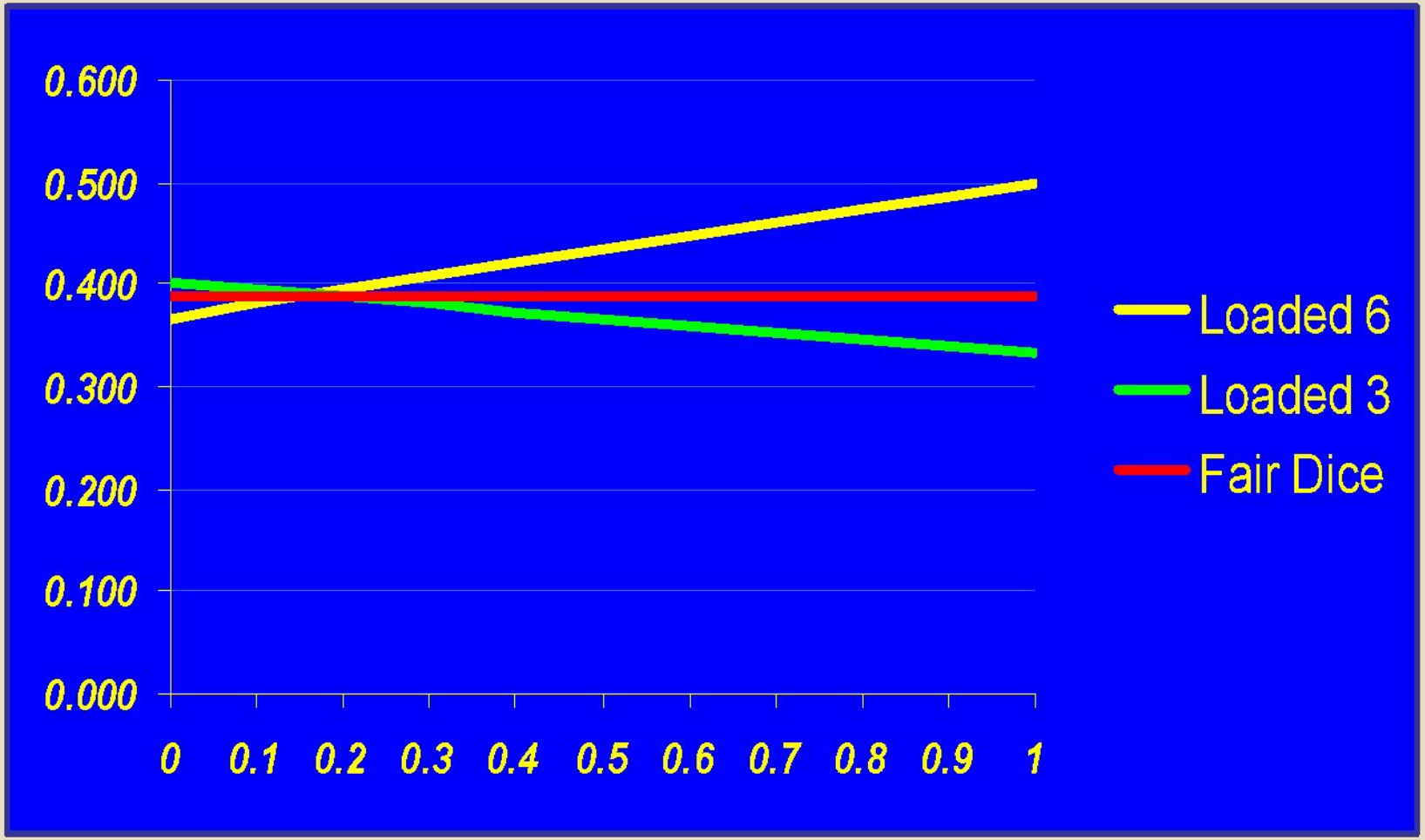
Proportion of Winning Hands

- With Fair Dice: 38.88%

- With One Loaded Die:
  - Lands 3, 25% of the time 38.3%
  - Lands 3, 35% of the time 37.7%

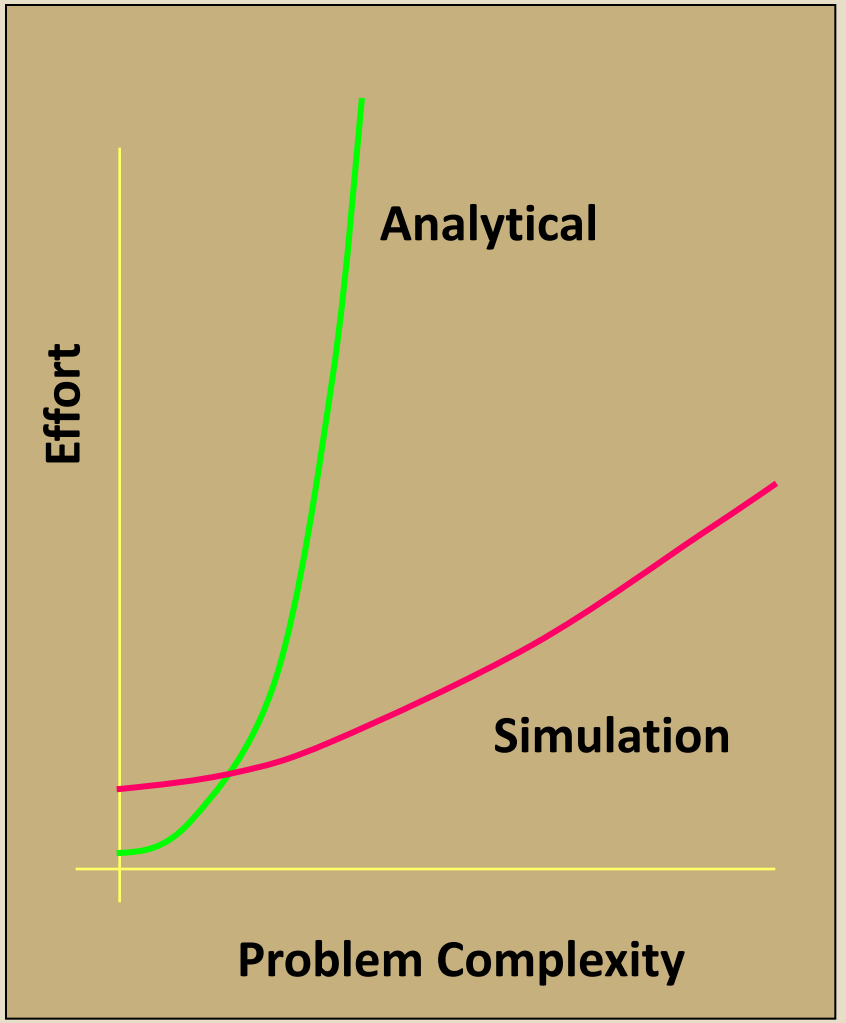
  - Lands 6, 25% of the time 40.0%
  - Lands 6, 35% of the time 41.3%

Analytical Solution



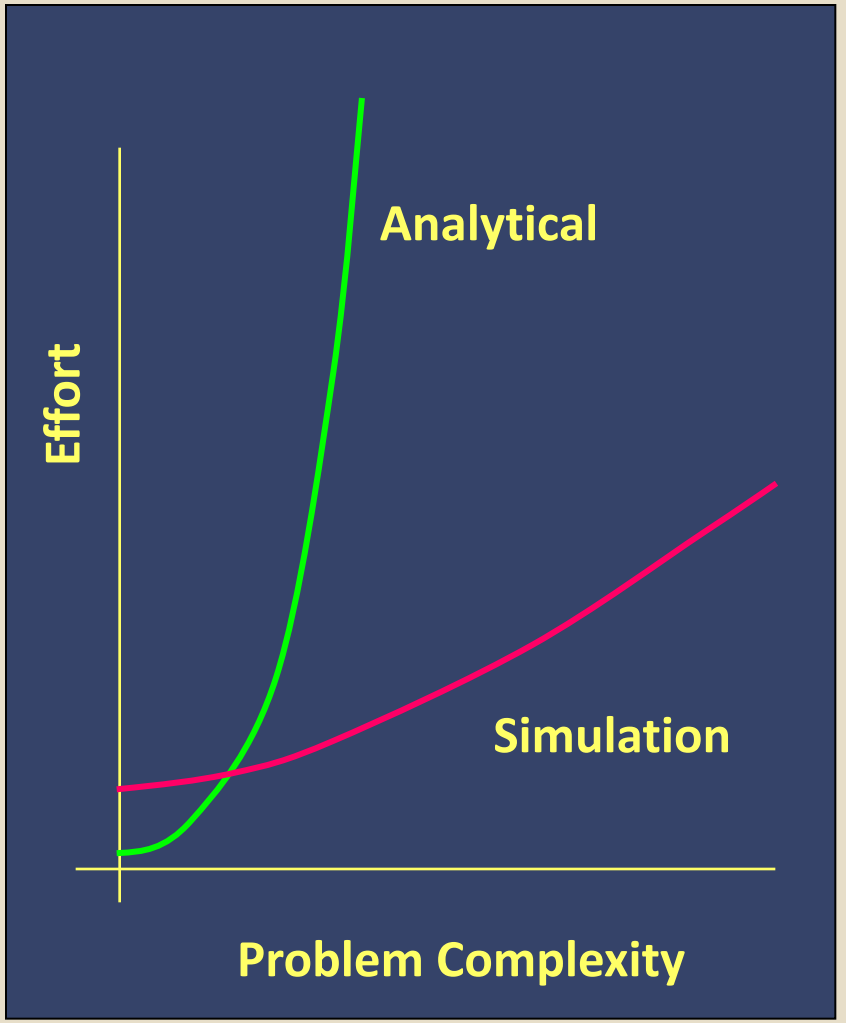
Analytical vs. Simulation

- Analytical solutions are exact, elegant, and defensible.
- But, they require enormous effort in real world problems.
- The required human resources are usually not available.



Analytical vs. Simulation

- Analytical solutions are exact, elegant, and defensible.
- But, they require enormous effort in real world problems.
- The required human resources are usually not available.



Monte Carlo Simulation Software

- Analytica™
  - Free version (for small to medium complexity models)
  - Professional Licence: https://analytica.com/products/free-edition/
- R™ Statistical Software Package
  - Free and unlimited
  - Often used with R Studio (free and commercial versions)
- @Risk™ (add-in to Microsoft Excel)
  - 15-day free trial: https://lumivero.com/resources/free-trial/atrisk/
  - Cost: USD $2125 per year

Quantitative Risk Assessment

A FEW MORE SIMPLE PROBABILITY DISTRIBUTIONS:

TRIANGULAR, PERT, BETA

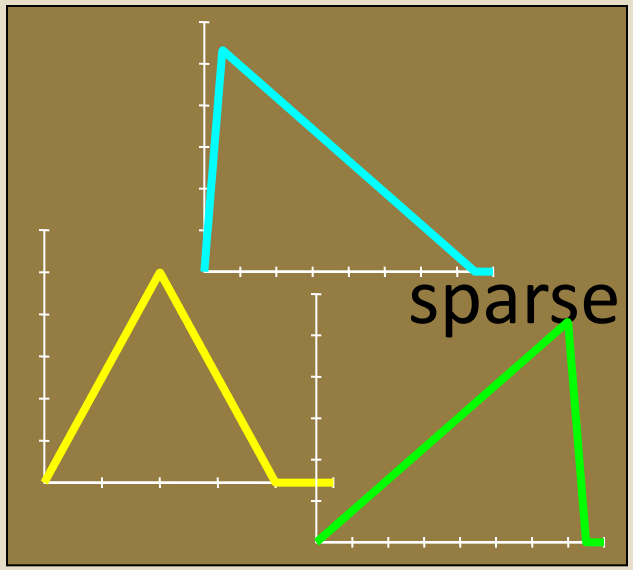
Distributions in Risk Assessment

- Many distributions are used in risk assessment modeling for public health
- We'll look at 3 more distributions now to give a quick sample of risk modeling

- Many statistics text books available on the subject
- Good resources:
  - https://en.wikipedia.org/wiki/Probability_distribution
  - https://mathworld.wolfram.com/topics/ProbabilityandStatistics.html

Triangular(a,b,c)

- Parameters: minimum (a); mode (b); maximum (c)
- Links the points
- Continuous
- Bounded
  - Domain:(a≤x≤b)
- **Mean = (a+b+c)/3**
- Often used when data are
  - “rough modeling”

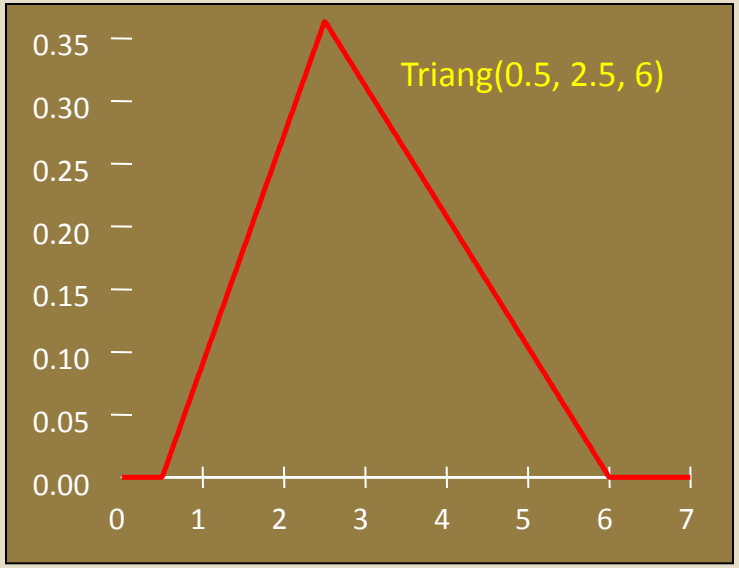


Example of Triangular Distribution

- A survey of patients shows that the most likely waiting time in the ER is 2.5 hours (with a minimum of 0.5 and a maximum of 6 hours)
- Triangular(0.5,2.5,6)
- **Mean= (a+b+c)/3**

  3 hours

  =RiskTriang(a,b,c)



Pert(a,b,c)

- Parameters: minimum (a); mode (b); maximum (c)
- Links points in a “bell”-like shape
- Continuous
- Bounded
  - Domain: (a≤x≤b)
- **Mean = (a+4*b+c)/6**
- Also used when data are sparse –
  - “rough modeling”
  - Similar, but smoother than triangular distribution.

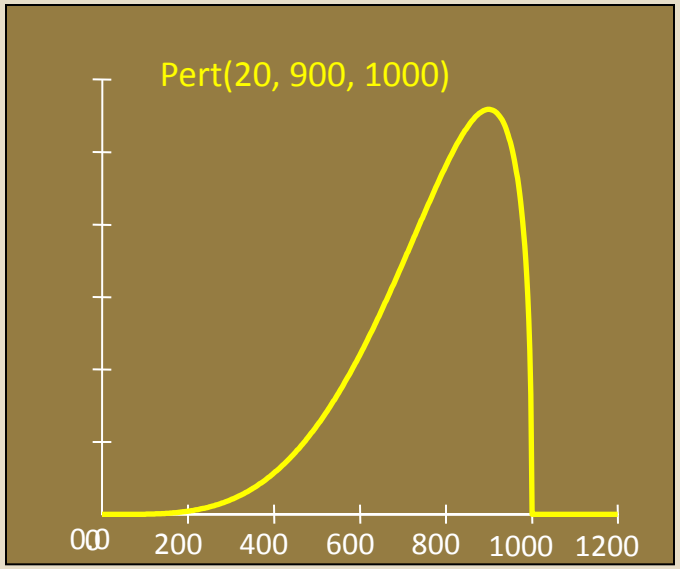
Example of Pert Distribution

- The mean concentration of salmonella in a contaminated raw egg is unknown. It is thought to have a minimum of 20cfu, a maximum of 1000 cfu and a most likely level of 900cfu. What is the level per egg?
- Pert (20,900,1000)
- **Mean = (a+4*b+c)/6**

  = (20+4*900+1000)/6

  = 770 cfu per egg

=RiskPert(a,b,c)



Beta(α, β)

- Describes probability of success (p) given s successes occurred in n trials
- Continuous
- Bounded
  - Domain: 0<x<1
- Mean = $\alpha/(\alpha+\beta)$
- Can be used to represent uncertainty in prevalence given test results with s positives and n-s negatives with no prior knowledge (α=s+1, β=n-s+1)

Example: Beta distribution

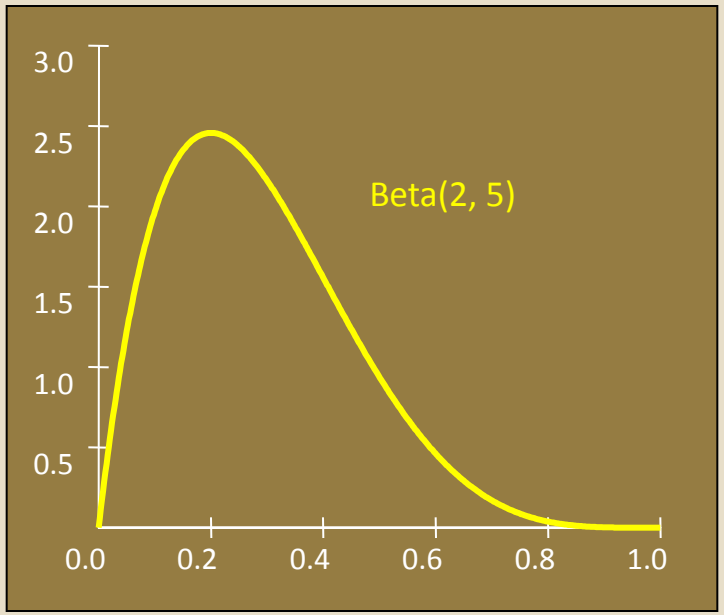
- Survey data show 1 positive and 4 negative blood tests (so ... s=1, n=5).
- Beta(1,1) represents the “ignorance distribution”
- We add the survey results to the ignorance distribution

- *Prev = Beta(s+1, n-s+1)*

  Beta(1+1,5-1+1)

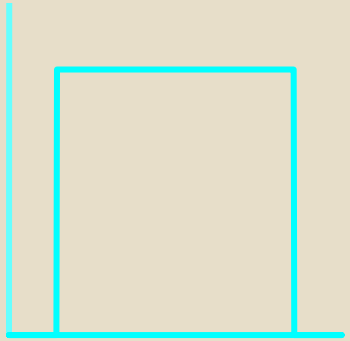
  Beta(2,5)

=RiskBeta(s+1,n-s+1)



Recap: 4 Distributions

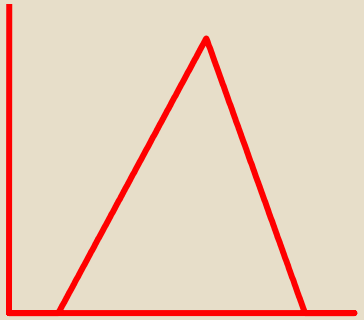
Uniform(min,max)



Beta(s+1,n-s+1)

Pert(min,ml,max)

Triangular(min,ml,max)



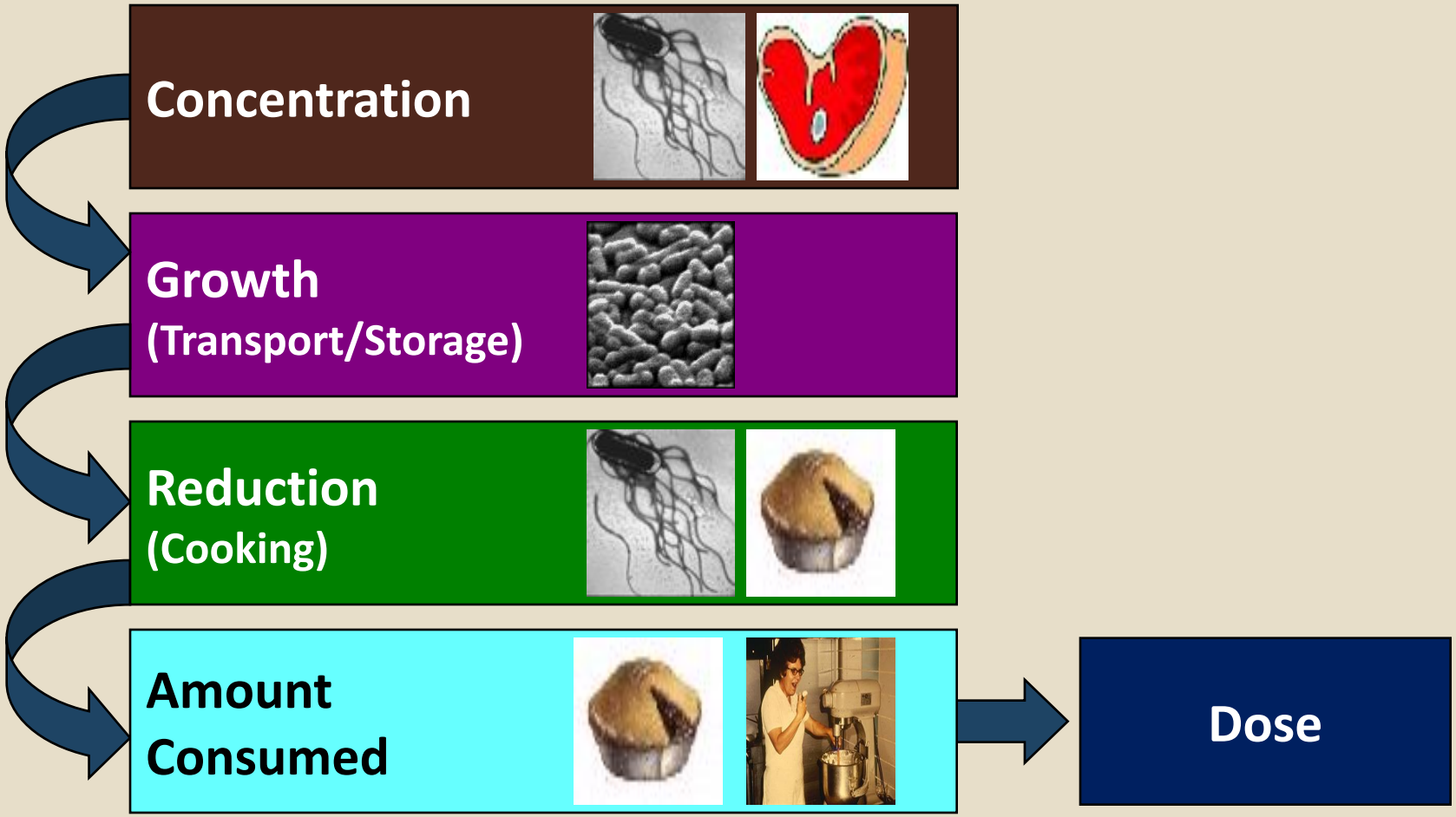
Quantitative Risk Assessment

COMPARING A DETERMINISTIC AND PROBABILISTIC APPROACH

Comparison of Deterministic and Probabilistic Solutions

- Remember the Meat Pies scenario
  - Deterministic model
- What if we included variation into the system?
- Question: Where might we want to include variation in the model?

Revisiting our Scenario



Example Scenario

Mean Values

- Bug “X” Concentration = 2.0 log CFU/g
- Bug “X” Growth = 1.5 log (unitless multiplier)
- Bug “X” Inactivation = 3.6 log (unitless mulitiplier)
- Serving Size = 53.33 g

Example Scenario

Worst Case (upper limit)

- Bug “X” Concentration = 4.0 log CFU/g
- Bug “X” Growth = 1.85 (unitless multiplier)
- Bug “X” Inactivation = 2.6 log (unitless multiplier)
- Serving Size = 85.00 g

Including Variability through Probability Distributions

Replacing Point Estimates with Distributions

- Bug "X" Concentration = Uniform (2.0, 4.0) log CFU/g
- Bug "X" Growth = Triangular (1, 1.5, 2) log change
- Bug "X" Inactivation = Triangular (2.5, 3, 5) log change
- Serving Size = Triangular (10, 50, 100) grams

Recall that for point estimates…

- If conservative point estimate falls below maximum acceptable risk, then we know that the risk is truly acceptable (Amount of overprotection is unknown)

- If conservative point estimate falls above maximum acceptable risk, then we do not know if the risk is truly unacceptable or result of propagated conservatism.

Burmaster 1995

Probabilistic vs. Point Estimate

- Using the mean value:
  - quite likely to occur - realistic
  - doses higher than this frequently occur - not conservative

- Using the conservative estimates
  - not very likely to occur - not realistic
  - doses higher than this rarely occur – “conservative”
  - Still, may not be conservative enough
    - Should 95% confidence be a surrogate for ‘safe’

Probabilistic vs. Point Estimate

- Point Estimates
  - Probability of an event occurring is not considered
  - Represents a significant loss of information.
  - Risk Management decisions made with very little information.
  - Assessments can be overly conservative, or inadequately protective, depending on the application.

Probabilistic vs. Point Estimate

- Selection of conservative estimate is a contentious issue:
  - How conservative should it be?
    - Worst Case Scenarios (creativity may the only limit to this)
    - Default regulatory guidelines
  - Propagating conservative estimates through assessment results in estimates of risk with no probability context
    - Reduces credibility of assessment
    - Risk Management decisions not “based on science“

Introduction to Model Analysis

- Review of Outputs from @Risk
  - Simulation files
  - Graphical Output and Reports
- Sensitivity Analysis (Excel and @Risk)
- Importance Analysis (@Risk)
- Running Multiple Scenarios

Quantitative Risk Assessment: Focus on Simulation and Exposure Assessment

UNDERSTANDING MONTE CARLO SIMULATION

Recall…

- ***Simulation*** - any analytical method meant to imitate a real-life system, especially when other analyses are too complex mathematically or too difficult to reproduce.

- For each probabilistic variable define the possible values with a distribution.

- Monte Carlo Analysis
  - systematically constructs the probability distribution of output variables, by randomly selecting values for input variables *according to their probability distributions*.

And that…

- The random selection process is repeated many times
  - multiple random scenarios, often referred to as “iterations”
- Some output values (combinations of inputs) are generated more often than others – this frequency distribution approaches the true probability distribution as the number of iterations increases (if we could actually know it analytically).

Generating distributions

- Based upon pseudo-random numbers
- Example method:

- Analytic inversion
- If u is uniformly distributed over (0,1), and Y has cumulative dist F_Y, then $F_Y^{-1}(u)$ has cdf F_Y
- Method
- Generate u, determine $x=F^{-1}(u)$, return x

Example

- Generate an exponential distribution
- Cdf is $F(x)=1-\exp^{(-\lambda x)}$
- Let $u= 1-\exp^{(-\lambda x)}$
- Therefore $x=-(1/\lambda)\ln(u)$
- Algorithm:
  - Generate u
  - Return $x=-(1/\lambda)\ln(u)$

Sampling methods: Monte Carlo

Simple Random Sampling (or, "Simple Monte Carlo")

- The most straightforward sampling method
- Samples U(0,1) with replacement
- Requires a relatively large sample size to generate accurate output statistics when complex models are simulated

Sampling Methods: Latin Hypercube

Latin Hypercube sampling

- Less common method (but more common in risk analysis due to availability in off-the-shelf software)
- Area under the distribution curve is segregated according to the sample size specified (referred to as iterations)
- Randomly samples once within each ‘area’

Convergence

- As the number of iterations increases, the statistics (mean, variance) of the simulated output distribution will converge toward the correct analytical solution.

- Accuracy of this answer can be improved by simulating more scenarios.

Optimal Settings

- There are no universally applicable procedures for determining the optimal simulation settings
- There are some general guidelines that a model developer may adopt to help ensure the number of iterations in the model simulation is of the appropriate magnitude

What is optimal?

- As the number of iterations increases the representation of the input and output distributions is improved
- When models include skewed distributions, highly non-linear equations or rare events the number of iterations required to achieve a good representation of the output distribution will be higher than models without these properties
- Aim is to determine point where “extra effort to achieve accuracy exceeds reward”

First Option

- Defining a “true mean” at a very large number of iterations, that is a number which is sufficiently beyond an expected convergence point of the model, and looking at the variation of the running mean from this “true mean” and accepting it if it is within some range
  - For example ±1%

Second Option

- Observing the change in the statistic over increasing iterations and accepting the results when the statistic no longer varies more than some acceptable level (for example ±1%).
  - No defined “true mean” rather the assumption is made that at the point where the stability is obtained represents the “true mean”.
  - Caution needed as a model can appear to stabilise but in subsequent iterations diverge wildly.
    - Non-linear (exponential, threshold) models and rare events.
  - Using the “true mean” as a criteria avoids this issue.

Steps to Perform

- Identify the output that represents a stable (or converged) model
- Identify the statistic(s) to monitor
- Define the criteria for stable (converged) estimates
- Run the model several times with different sample sizes (numbers of iterations)

Steps to Perform (continued)

- Examine the results and determine the point where the model convergence is acceptable
- The convergence of the model should be re-examined whenever there is a change to the model in either the distributions used to describe the variables in the model, or changes in the model equations themselves.

- Test is if multiple runs give approximately same results
  - Different number seeds
  - Allowing for randomness!

Quantitative Risk Assessment: Focus on Simulation and Exposure Assessment

INSIGHT INTO COMMONLY-USED PROBABILITY DISTRIBUTIONS

Process-derived distributions

- Shape of the distribution comes from the mathematics describing a theoretical phenomenon
  - Also referred to as ‘mechanistic’
- Requires an understanding of the underlying random process, and any randomness that is part of observing it.
- Theoretical basis for a particular distribution may be used to ‘overrule’ goodness of fit statistics that would suggest other distributions appear to be preferable.

‘Empirical’ distributions

- Used for unknown underlying process, mixed data, or when the phenomenon is too complex to assign to a theoretical class
- Used to capture subjective judgments (e.g. prior beliefs, expert beliefs)
- Also, applies to distributions based directly on data, regardless of process

Process-derived vs. Empirical

Process-derived	Empirical
• Binomial,Negative Binomial • Exponential • Gamma * • Geometric,Hypergeometric • Gumbel • Normal, Lognormal • Poisson • Weibull	• Beta *, Beta-Pert ** • Uniform * • Triangular • Empirical PDF, CDF based directly on data

* Commonly play formal roles in Bayesian Updating

** Often used for expert judgment, or first guesses

Another way of classifying distributions

- Unbounded
  - extends from negative to positive infinity
- Partially bounded
  - constrained at one extreme (often zero)
- Two-sided bounded
  - confined within lower and upper bounds
- Fixed domain
  - Can take on a fixed number of values

THE NORMAL AND LOGNORMAL DISTRIBUTIONS AND THE CENTRAL LIMIT THEOREM

The Normal distribution

- Bell-shaped symmetrical curve
- Normal(μ,σ)
- μ is mean and σ is standard deviation
- Continuous
- Unbounded
  - Domain: -inf<x<inf
  - (Kurtosis = 3)

The Log-Normal (Lognormal) distribution

- Related to the normal distribution:
  - When the log-normally distributed data are log-transformed, they follow a normal distribution
- Various parameterizations are used (be careful here):
  - LogNormal(logμ, logσ)
  - LogNormal(median, gsd)
- Continuous
- Bounded from below
  - Domain: 0<x<inf

Central Limit Theorem

- Simply put, the distribution of the sum of a sufficiently large number of independent random variables will converge toward the normal distribution as the number of variables increases.

Product of RVs Lognormal

- What about the *product* of a sufficiently large number of random variables?

  Remembering that:

  If ProdX = X1 * X2 * X3* … Xn, then

  Log(ProdX) = Log(X1) + Log(X2) + Log(X3) + … Log(Xn)

- Since a product can always be re-written as a sum of log-transformed random variables, the CLT predicts that the log of this product will be normally distributed.
  - Therefore the product must be log-normally distributed.

Central Limit Theorem Example

- Take the sum of n random variables, each distributed as a uniform distribution between 0 and 1: Uniform(0,1)
- Let's look at the distribution of the mean with increasing n

2

4

6

8

10

.

.

.

.

50

Sum of 50 U(0,1)

Kurtosis ≈ 3

IMPORTANT PROCESS-DERIVED DISTRIBUTIONS FOR CHARACTERIZING PHYSICAL SYSTEMS AND THEIR OBSERVATION

Parametric distributions

Three key discrete random processes used in risk assessment are:

1. Binomial (with Beta)
2. Poisson (with Exponential and Gamma)
3. Hypergeometric (less common)

Binomial (Bernoulli) Process

- Given a number of independent trials (n)
- Two possible outcomes of each trial - success or failure
- A Binomial random variable counts the number of successes (s) among the n trials.
- Probability of success = p
- Probability of failure =1-p

Binomial Process!

Examples of Binomial Process

- The obvious one….flipping a coin!
  - Toss coin n times will get a head s times

  (considering a head as a success!)
  - There is a probability p of getting a head
  - There is a probability 1-p of getting a tail

- Picking people from a crowd - will either be male or female
- Number of animals with “Disease X” selected from a herd - either diseased or not

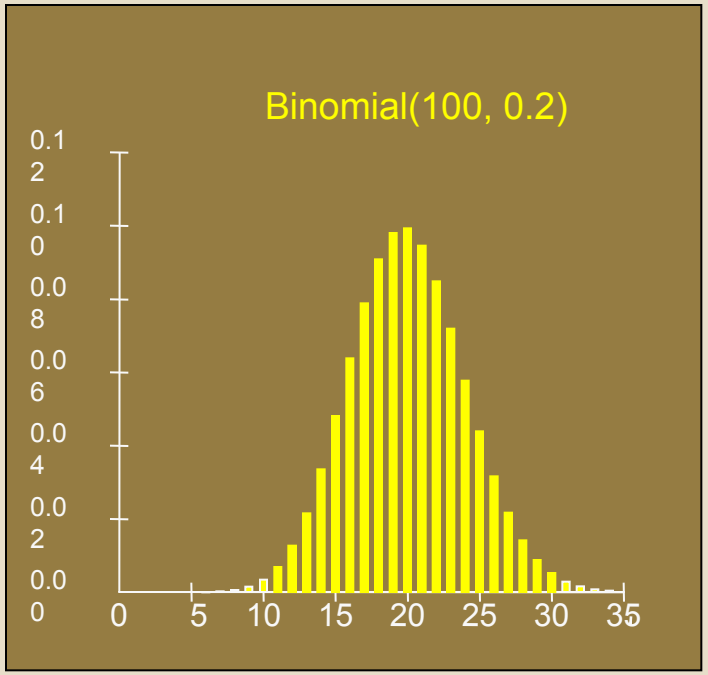
Binomial(n,p)

- Describes number of successes (s) given n trials, each with probability of p for success
- Discrete
- Bounded
  - Domain: (0≤x≤n)
- **Mean = np**

- **Bernoulli is a special case of Binomial with n=1**

Binomial - example

- TV switches have 0.2 probability of being faulty. How many are faulty in a random batch of 100?
  - Binomial(100,0.2)
  - Mean=20



Binomial calculations

- The probability that a person is allergic to cats is 0.3. What is the probability of at least one in a group of 50 people, selected at random, being allergic?

- Probability that a person is not allergic
  - $(1-p) = (1-0.3)$ or 0.7
- Probability all persons in group not allergic
  - $(1-p)^n = (1-0.3)^{50}$ or 0.7^{50}
- Probability at least one person in group is allergic
  - $1- (1-p)^n =1- (1-0.3)^{50} = 0.999999982$

Beta(α, β)

- Revisiting the Beta distribution
- We talked about Beta(s+1,n-s+1)
- But it really should be thought of as:
- Beta(α, β)
  - α =s+a
  - β =n-s+b
  - a and b depend on prior, where prior is Beta(a,b).

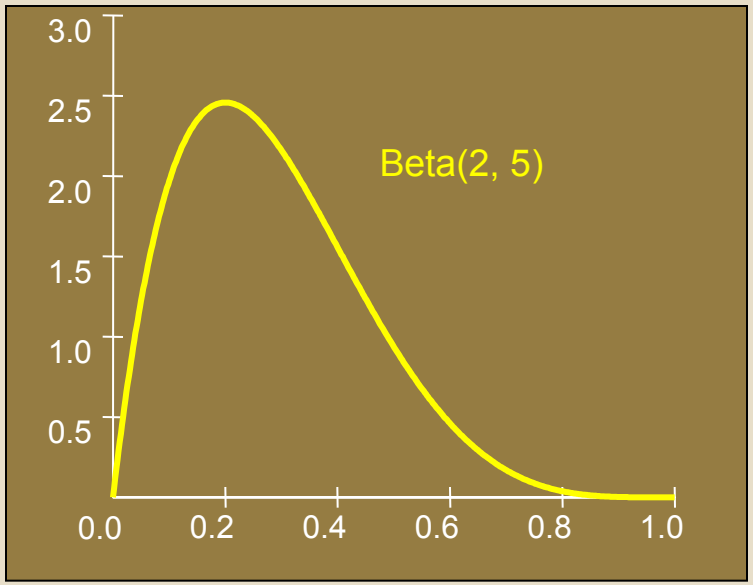
Often the prior knowledge is 'ignorance' which is reflected by Beta(1,1) which yields α=s+1 β=n-s+1 after observing n trials.

Updating Beta with New Information

- *Start with Beta(1,1) prior i.e. ignorance*
- Survey data show 1 positive and 4 negative blood tests, add them to the ignorance distribution to get

- *Prev = Beta(s+a, n-s+b)*

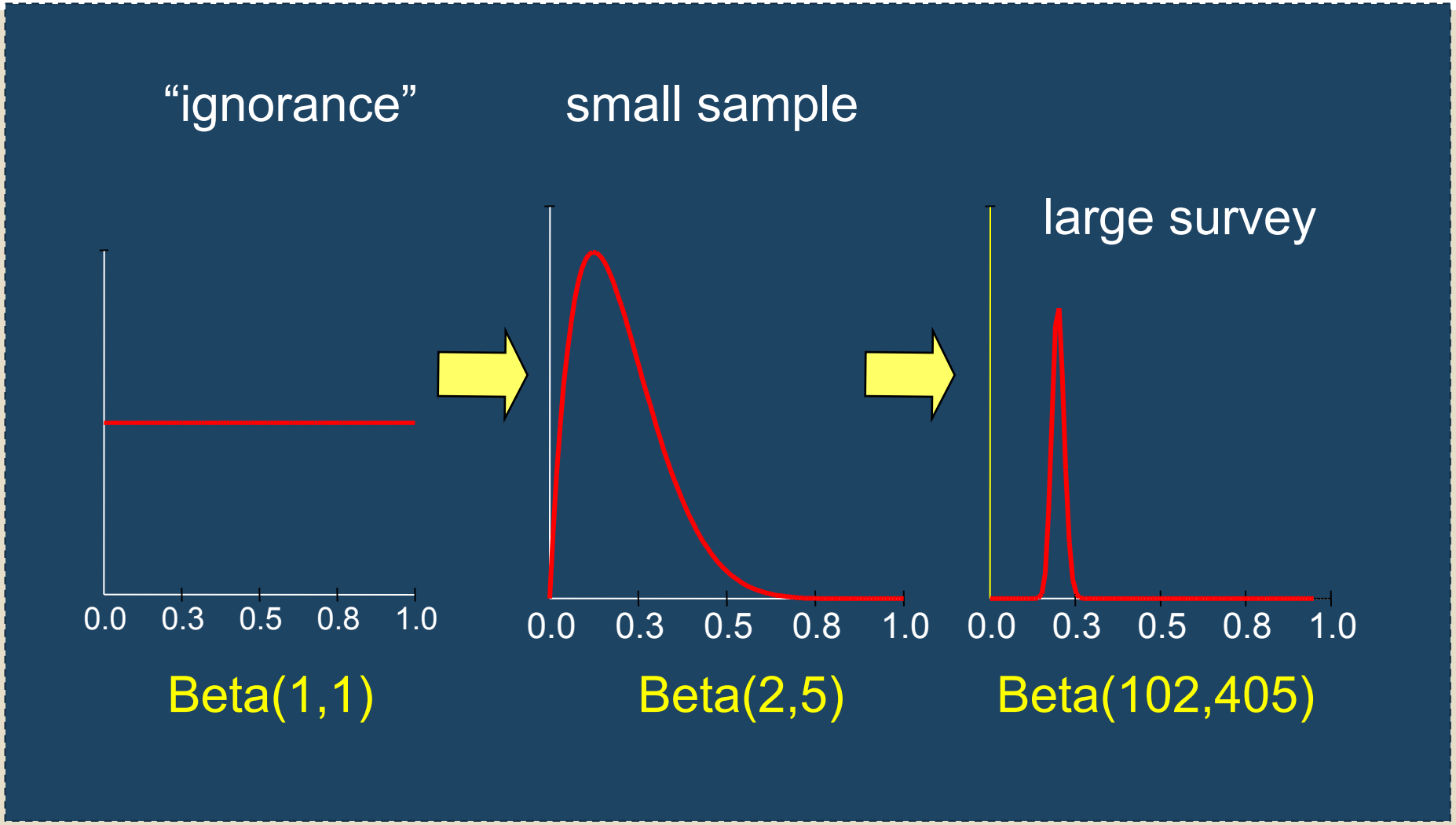
  Beta(2,5)



Beta: More Data Leads to Tighter Distributions

- If another test was done on 500 farms, and 100 were positive.
  - Remember Beta(s+a, n-s+b)
- From previous test a=2, b=5
- Therefore an appropriate uncertainty distribution would be Beta(102,405).
  - Mean is still ~20%, but distribution is narrowly distributed, reflecting increased confidence.

‘Bayesian Updating’

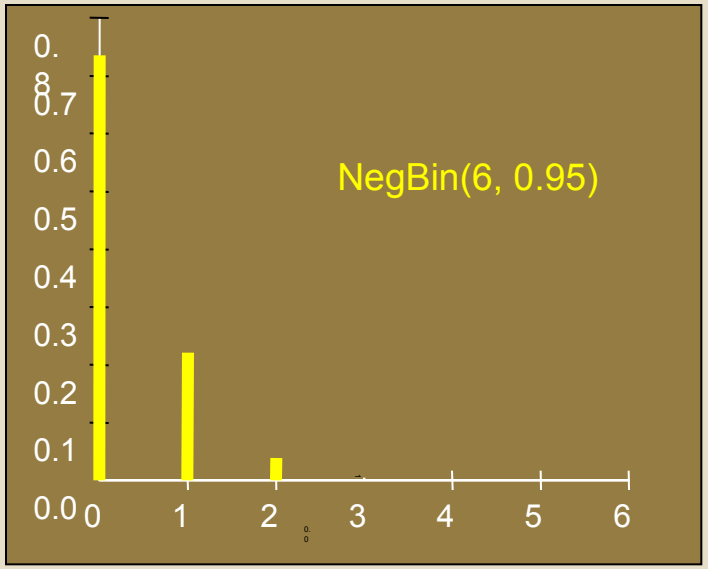


Negative Binomial(s,p)

- Describes the number of failures in a discrete process with success probability p, until s successes, each process stops at last success
- Discrete
- Bounded at 0
  - Domain: {0,1,2,3,…}
- Can also be used to reflect ‘over-dispersion’ in a Poisson process (discussed later).

Negative Binomial - Example

- Patients tested using assay with sensitivity of 95%, testing stopped at 6th positive result
- How many positives are we likely to have misdiagnosed? NegBin(6,0.95)
  - Mean = 0.315
  - P(missed at least one) = 36.5%



Poisson process

- A Poisson process is one in which events happen randomly within some window of opportunity.
- The Poisson distribution counts the number of observations of the process in a certain window.
- Occurs over a continuum of opportunity
- Observations of the process described by **λ**
  - **Observation process depends on both the intensity of the process and the extent of observation (time, distance, etc.)**

The Poisson Parameter (λ)

- A Poisson process has a single parameter, **λ**

  where **λ = μ w**

- **μ** is the average intensity of the process
  - This may be over time or space or other unit of measurement
  - E.g. # events per unit time, or # events per unit space
- "**w**" is the size of the window of observation
  - Units of time, area, volume etc.

Poisson Process

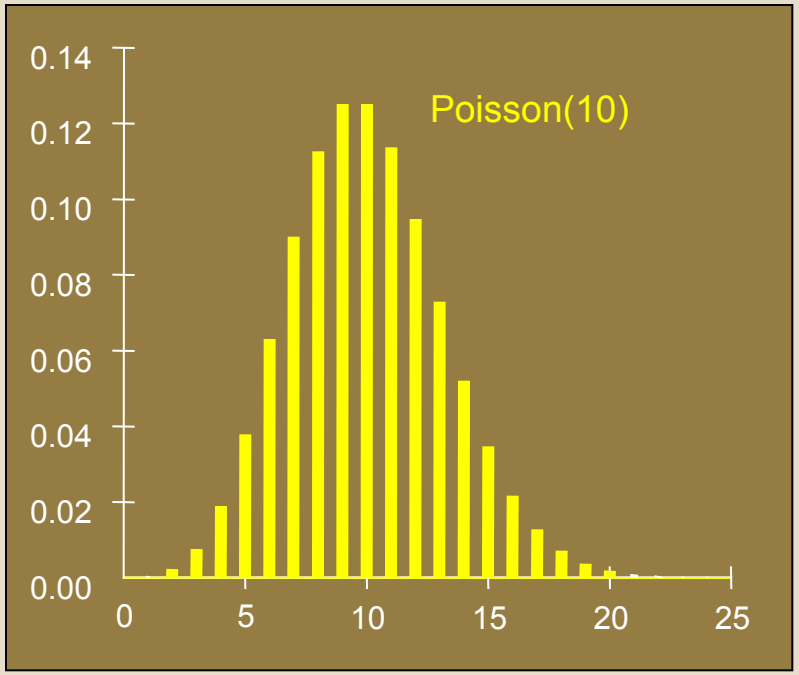
- 3 main distributions

1. Poisson (λ)
2. Gamma (α,β)
3. Exponential (β)

Poisson(λ)

- Describes the number of events (α) that occur given μw (i.e. λ)

- Discrete
- Bounded at 0
  - Domain: {0,1,2,…}
- **Mean = λ**



Poisson - Example

- Accidents occur at an average of 1 per 100 kilometers per year (μ=0.01 per kilometer-year).

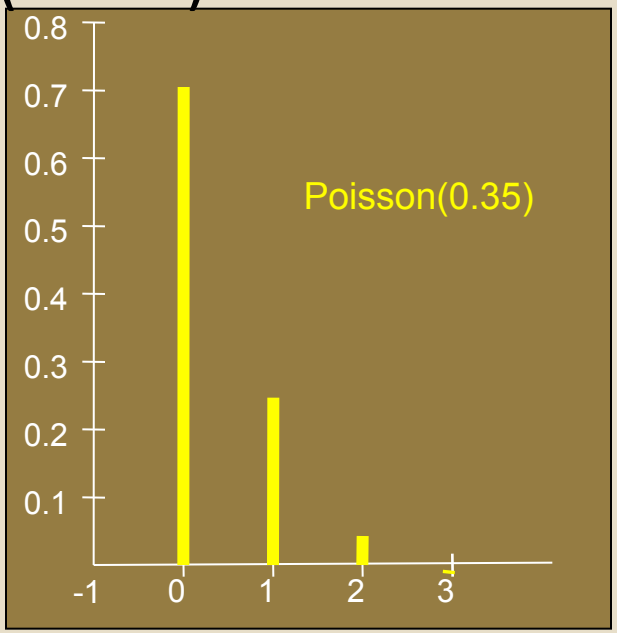
- If we observe a 1000 km stretch for one year (w = 1000*1), we can model the distribution of the number of accidents per year observed as:

  RiskPoisson(0.01*1000) or
  RiskPoisson(10)

Poisson – Example (2)

- A pathogen is randomly distributed throughout a homogeneous food product. Concentration is thought to be 1 CFU per 100 g (μ=0.01 CFU/g). A consumer eats 35 g of the product (w=35)

- The ingested dose can be modeled as:
  RiskPoisson(0.01x35) or
  RiskPoisson(0.35)



Poisson Calculations

- Probability of any particular count:

$$P(x) = \lambda^x e^{-\lambda} / x!$$

$$x! = x(x-1)(x-2)\ldots(2)(1)$$

- Important results:
  - Probability of zero observations:

$$P(x=0) = e^{-\lambda}$$

  - Probability of at least one:

$$P(x>0) = 1 - e^{-\lambda}$$

Example

- In country Y the mean annual number of cases of Creutzfeldt-Jakob Disease is 7.

- What is the probability there will be 0 cases next year?
  - $P(x=0) = e^{-\lambda} = e^{-7} = 0.000912$
- What is the probability there will be at least one?
  - $P(x>0) = 1 - e^{-\lambda} = 1-0.000912 = 0.999088$

APPROXIMATIONS OF ONE DISTRIBUTION BY ANOTHER

Approximations

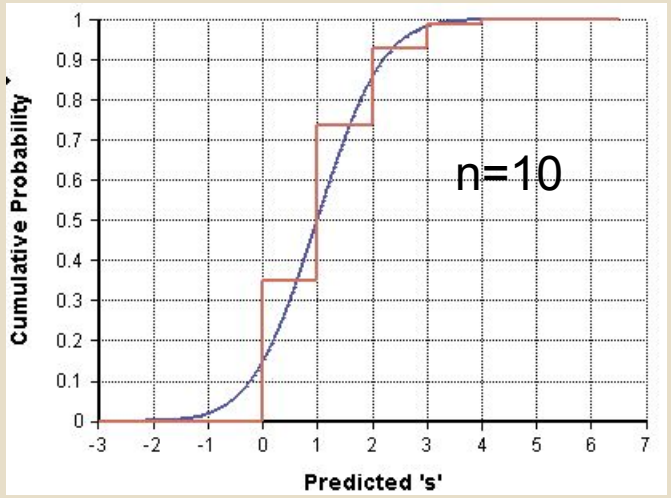
- In some circumstances it is convenient to use approximations to distributions
  - If a situation requires calculation of large numbers, or factorials of large numbers
    - Binomial or Poisson distribution
- Approximations can be applied given certain conditions are met

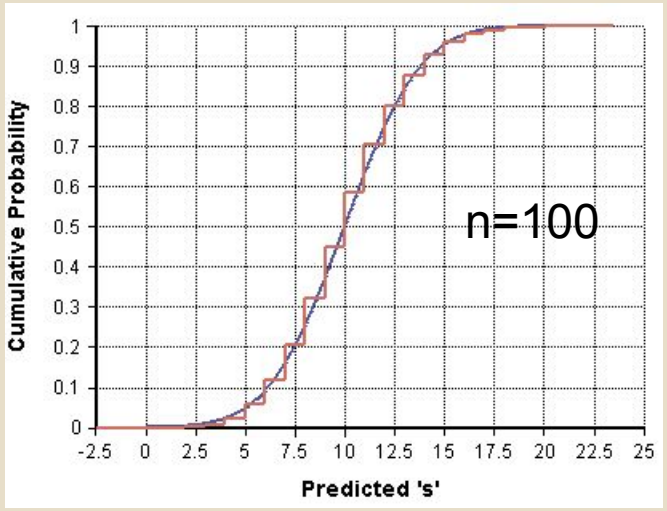
Binomial Distribution

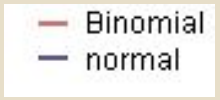
- Pmf is given by $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Involves calculation of factorials
- What if I toss a coin 1 million times…requires calculation of factorials up to 1 million!
- Can be approximated by the Normal distribution

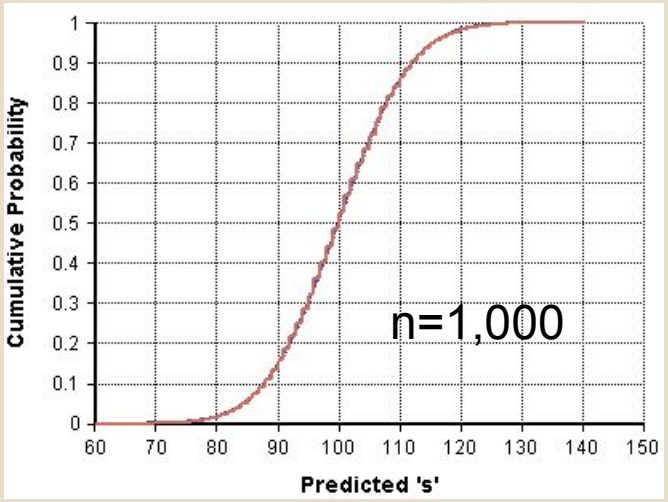
- Binomial(n,p) ≈ Normal(np,$(npq)^{0.5}$)
  - q=(1-p)
  - One possible criterion for use $n^{0.31}p>0.47$

Normal Approximation to Binomial, p=0.1









Quantitative Risk Assessment Methods

MODELING UNCERTAINTY AND VARIABILITY (ADVANCED CONCEPTS IN SIMULATION)

Uncertainty

- Uncertainty is used to describe the fact that we have incomplete knowledge.

- Uncertainty can be treated:
  - formally (e.g. sampling error)
  - quasi-formally (e.g. formal expert elicitation)
  - informally (e.g. judgement)

Variability

- Variability refers to the fact that natural phenomena have inherent dispersion.
- This type of dispersion is not reducible through sampling or research
- Reduction of dispersion is not an improvement in knowledge…
  it would reflect a loss of information.

Uncertainty and Variability

- Imagine you measure the height of 10% of a university class
- Data represent variability
- But not a complete sample, so also have uncertainty
- More you sample – less uncertainty still
- Sample 100% and you have perfect knowledge of the distribution of variability
- But what if you can't increase the sampling?.....

Examples of Uncertainty & Variability

Uncertainty	Variability
○ Effectiveness of boiling eggs for 7 min. ○ Probability of any single farm being positive ○ Proportion of consumers who eat product raw	○ Duration of boiling of eggs by consumers ○ Variation in size of herds ○ Size of portion consumed

One (imperfect) way to differentiate

- Construct the following statement regarding a distributed parameter, P:
  - "With perfect information, P could be reduced to a single value."
- If it sounds plausible … Uncertainty
- If it sounds inappropriate … Variability
- Most phenomena are modelled with both U & V
  - Often our most important uncertainty is the extent of variability

Two-stage (or 2-D) Monte Carlo

- It is often advocated to separate uncertainty and variability.
  - Conceptually, it makes sense to differentiate measures of ignorance and measures of real variability.
  - Practically, it is very difficult to do completely.
- Two stages of simulation
  - Simulate values for uncertain random variables
  - Use uncertain random variables to drive a series of simulations which explore only variability.
  - Analyze the variability and uncertainty separately.
- FDA-iRISK fully supports 2-D Monte Carlo simulation
- R™ package mc2d also supports 2-D Monte Carlo simulation