

Stochastic disease models

“Real life can be truly chaotic.”

Taylor, 2003

A brief review

- The epidemic model developed by Reed and Frost represents a class of mathematical models called chain binomial models
 - We have not yet discussed what a “chain binomial model” is
- The Reed-Frost model operates in discrete time units, where each time period is equal to the length of the average serial interval (\approx average incubation period) for the disease being modeled
- The number of cases in a particular time period can be calculated based on the number of cases from an earlier time period
- This calculation also uses the average number of adequate contacts (designated k) that each individual has with others in the population during a single time period

The Reed-Frost equation (more review)

- $C_{t+1} = S_t(1 - q^{C_t})$

where:

t indicates the time period (*constant duration*)

C = # of cases (infectious individuals)

S = # of susceptible individuals (*constant*)

q = prob. of avoiding adequate contact (*constant*)

- $q = 1 - (\text{prob. of adequate contact}) = 1 - p$

- $p = k / (N - 1)$

where:

k = average number of adequate contacts by an individual in a single time period (*constant*)

N = size of the population (*constant*)

A closer look at k (I)

- The number of adequate contacts that an individual has in a time period could be calculated as follows:
 - (# of *all* contacts in a time period)
 - × (probability that a contact – if it occurs – will be adequate)
 - Note that this probability does not represent the same thing as p

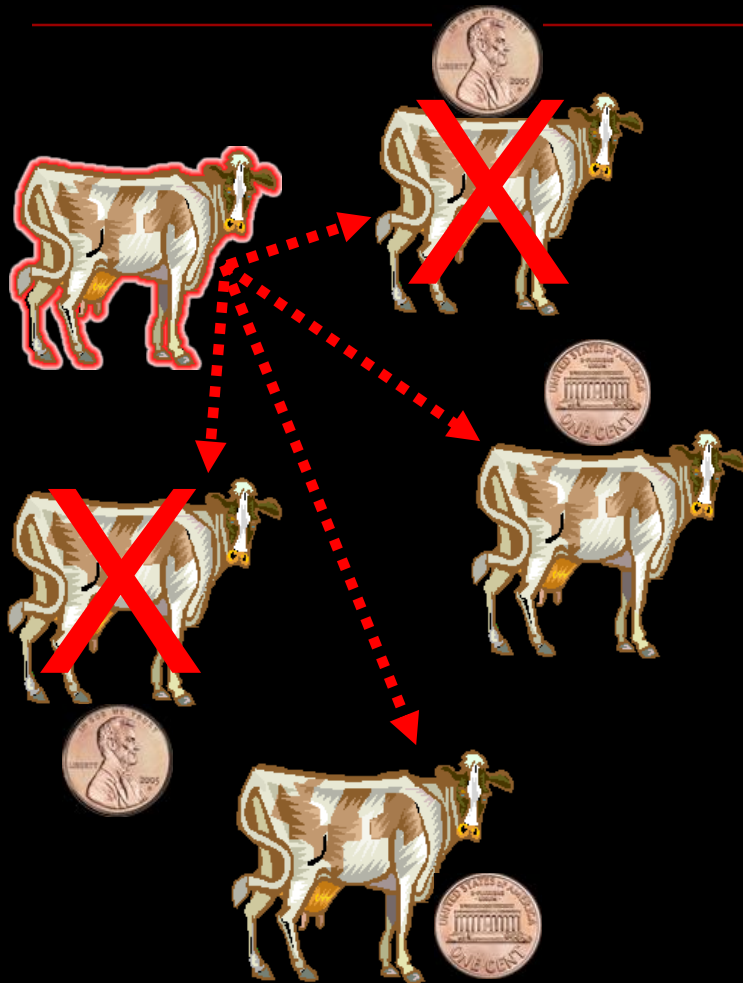
- Let's denote this symbolically as:

$$k = k_A \times h$$

- Now let's consider an example in which every individual will have a total of exactly four contacts per period, and in which each contact that occurs has a 50% chance of being adequate:

$$k_A = 4, h = 0.5, k = 4 \times 0.5 = 2$$

A closer look at k (II)



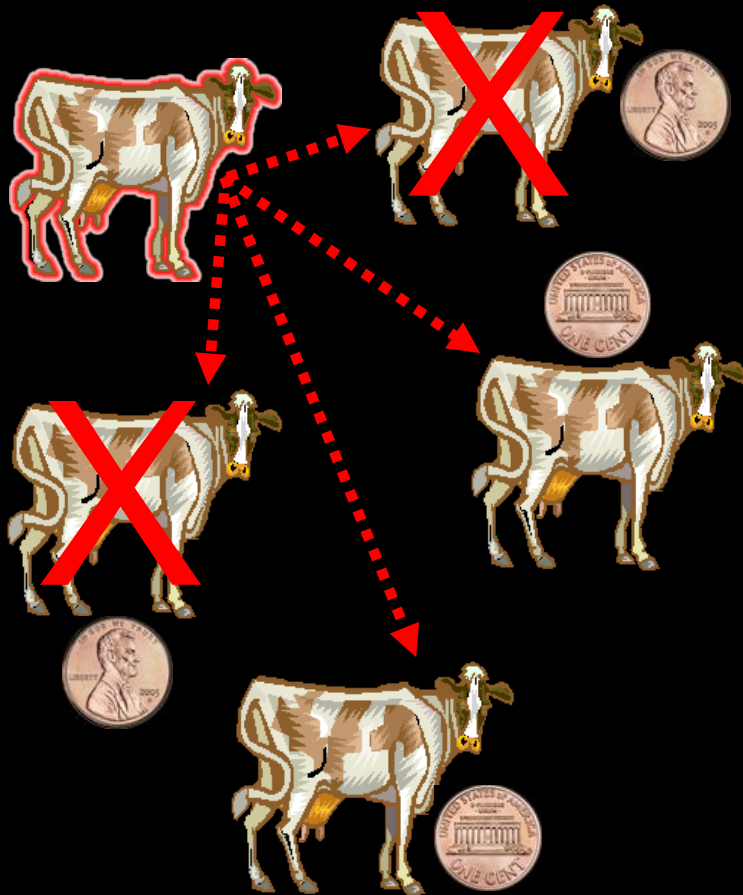
Based on this example:

- There will be 4 contacts
- Each contact has a 50% chance of being adequate
- All recipients are susceptible, so every adequate contact will be effective
- We expect that there will be 2 infected individuals in the next time period
- The deterministic model we've used so far will always produce this result
- But this is not the only possible result

A closer look at k (III)

- Of the two components of k ...
 - k_A , the total number of contacts
 - h , the probability that a contact will be adequate
- ...we could allow k to vary by incorporating stochasticity for the total number of contacts, and/or the probability that a contact will be adequate
- Let's consider the latter situation first...

Adequate contact as a binomial process



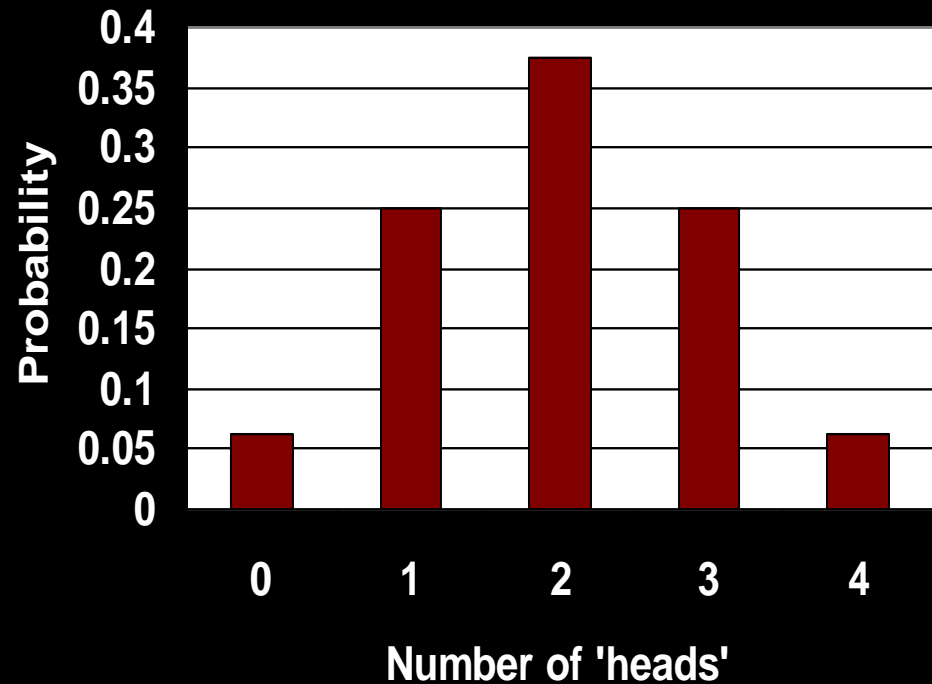
- Contact could be modeled as a process with two possible outcomes
 - Either contact happens, or it doesn't
- A contact has a certain probability of being adequate
 - In our example, this probability is 50%
 - 50% is an arbitrary choice: the two outcomes do not have to have an equal probability

A simple binomial process



etc.

Probabilities of n 'heads' in four coin tosses

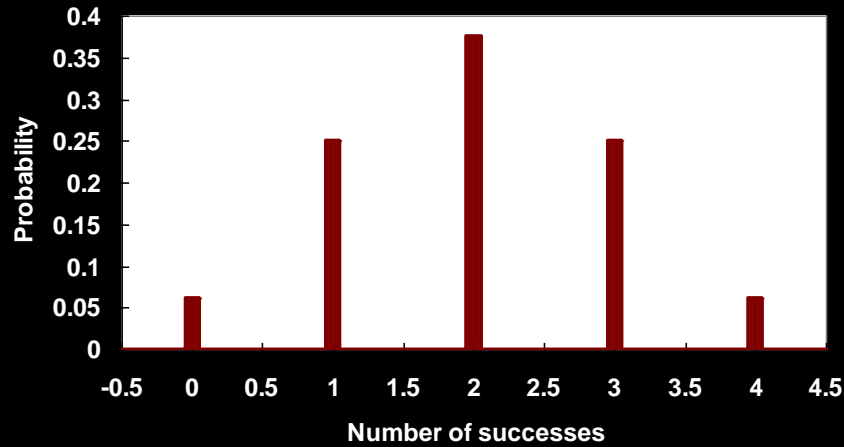


Characteristics of binomial processes

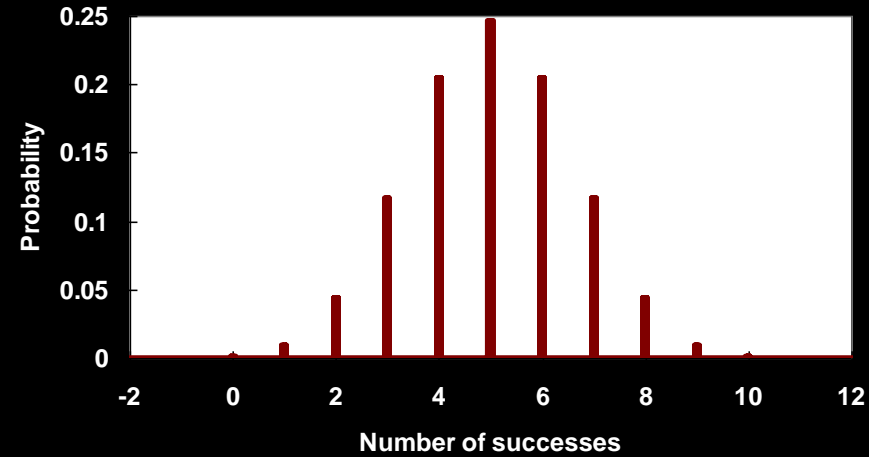
- A binomial process is one in which:
 - Each *trial* (coin toss; contact) has two possible outcomes (heads or tails; infection or no infection)
 - The outcome event can be defined as a success (heads; infection) or a failure (tails; no infection)
 - Each trial has the same probability of success (0.5 probability of heads; probability of infection of h)
 - All trials are independent of one another (the result of one coin toss has no influence on the result of the next toss)
- The number of successful events that might occur in some number of trials follows a *binomial distribution*

Some binomial distributions

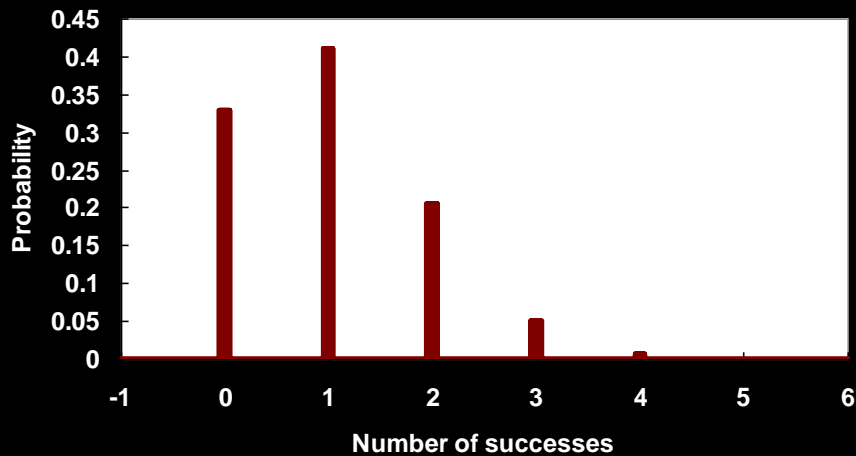
Binomial(4, 0.5)



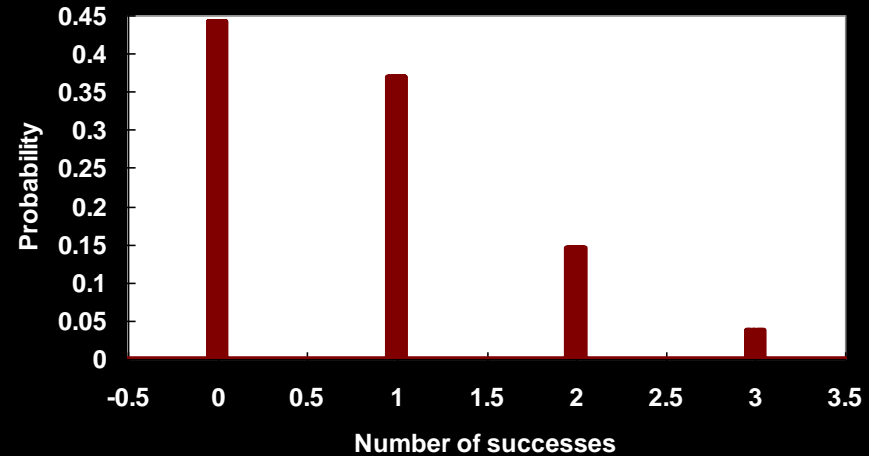
Binomial(10, 0.5)



Binomial(5, 0.2)



Binomial(20, 0.04)



Calculating a binomial probability

- To calculate the probability of x successes in n trials, where each trial has the probability of success p :

$$\frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$$

- For example: what is the probability of getting exactly 2 heads in 5 tosses of a fair coin?
 - $n = 5, x = 2, p = 0.5$

$$\frac{(5 \times 4 \times 3 \times 2 \times 1)}{(2 \times 1)(3 \times 2 \times 1)} (0.5)^2 (1-0.5)^3 = \frac{120}{(2 \times 6)} (0.313) = 0.3125$$

- In Excel, use the following formula: =BINOMDIST(2, 5, 0.5, false)

An aside: The origin of the term “chain binomial”

- The (deterministic) Reed-Frost formula, once again:
$$C_{t+1} = S_t (1 - q^{C_t})$$
- The probability of having exactly C_{t+1} cases in time period $t+1$ can be calculated with the binomial function¹, *i.e.*, what is the probability of C_{t+1} successes in S_t trials, where the probability of success is $(1 - q^{C_t})$?

$$P(C_{t+1}) = \frac{S_t!}{C_{t+1}!(S_t - C_{t+1})!} (1 - q^{C_t})^{C_{t+1}} (q^{C_t})^{(S_t - C_{t+1})}$$

- A similar calculation could be performed for each time period, resulting in a sequence or *chain* of binomial functions

¹Abbey, 1952

So, where were we?

- Let's modify our basic Reed-Frost model so that k varies for each time period, according to a binomial distribution
- k is no longer constant
- Consequently, neither will p or q
- The result will be an epidemic that is *possible* under the conditions we've specified
- It probably won't be the "average" epidemic

Results of a Reed-Frost model where k is not constant: Calculating p and q

| Time | Cases | Immune | Susceptible | k | p | q |
|------|-------|--------|-------------|-----|------|------|
| 0 | 1 | 0 | 100 | 2 | 0.02 | 0.98 |
| 1 | 2 | 1 | 98 | 2 | 0.02 | 0.98 |
| 2 | 4 | 3 | 94 | 4 | 0.04 | 0.96 |
| 3 | 14 | 7 | 80 | 3 | 0.03 | 0.97 |
| 4 | 27 | 21 | 53 | 1 | 0.01 | 0.99 |
| 5 | 13 | 48 | 40 | 2 | 0.02 | 0.98 |
| 6 | 9 | 61 | 31 | 3 | 0.03 | 0.97 |
| 7 | 7 | 70 | 24 | 1 | 0.01 | 0.99 |
| 8 | 2 | 78 | 21 | 3 | 0.03 | 0.97 |
| 9 | 1 | 79 | 21 | 1 | 0.01 | 0.99 |
| 10 | 0 | 80 | 21 | 1 | 0.01 | 0.99 |

- Like k , p , and q vary for each time period
- $p = k/(N-1)$, but k is different at each step

Model parameters:

- Population size: $N = 101$
- Initial number of cases: $C_0 = 1$
- Initial number of susceptibles: $S_0 = 100$
- $k \in \text{Binomial}(4, 0.5)$

Results of a Reed-Frost model where k is not constant: Calculating the number of new cases

| Time | Cases | Immune | Susceptible | | k | p | q |
|------|-------|--------|-------------|--|-----|------|------|
| 0 | 1 | 0 | 100 | | 2 | 0.02 | 0.98 |
| 1 | 2 | 1 | 98 | | 2 | 0.02 | 0.98 |
| 2 | 4 | 3 | 94 | | 4 | 0.04 | 0.96 |
| 3 | 14 | 7 | 80 | | 3 | 0.03 | 0.97 |
| 4 | 27 | 21 | 53 | | 1 | 0.01 | 0.99 |
| 5 | 13 | 48 | 40 | | 2 | 0.02 | 0.98 |
| 6 | 9 | 61 | 31 | | 3 | 0.03 | 0.97 |
| 7 | 7 | 70 | 24 | | 1 | 0.01 | 0.99 |
| 8 | 2 | 78 | 21 | | 3 | 0.03 | 0.97 |
| 9 | 1 | 79 | 21 | | 1 | 0.01 | 0.99 |
| 10 | 0 | 80 | 21 | | 1 | 0.01 | 0.99 |

- The number of cases depends on a different q value for each time period

Model parameters:

- Population size: $N = 101$
- Initial number of susceptibles: $S_0 = 100$
- Initial number of cases: $C_0 = 1$
- $k \in \text{Binomial}(4, 0.5)$

Results of a Reed-Frost model where k is not constant (I)

| <i>Time</i> | <i>Cases</i> | <i>Immune</i> | <i>Susceptible</i> | | <i>k</i> | <i>p</i> | <i>q</i> |
|--------------|--------------|---------------|--------------------|--|----------|----------|----------|
| 0 | 1 | 0 | 100 | | 2 | 0.02 | 0.98 |
| 1 | 2 | 1 | 98 | | 2 | 0.02 | 0.98 |
| 2 | 4 | 3 | 94 | | 4 | 0.04 | 0.96 |
| 3 | 14 | 7 | 80 | | 3 | 0.03 | 0.97 |
| 4 | 27 | 21 | 53 | | 1 | 0.01 | 0.99 |
| 5 | 13 | 48 | 40 | | 2 | 0.02 | 0.98 |
| 6 | 9 | 61 | 31 | | 3 | 0.03 | 0.97 |
| 7 | 7 | 70 | 24 | | 1 | 0.01 | 0.99 |
| 8 | 2 | 78 | 21 | | 3 | 0.03 | 0.97 |
| 9 | 1 | 79 | 21 | | 1 | 0.01 | 0.99 |
| 10 | 0 | 80 | 21 | | 1 | 0.01 | 0.99 |
| <i>TOTAL</i> | <i>80</i> | | | | | | |

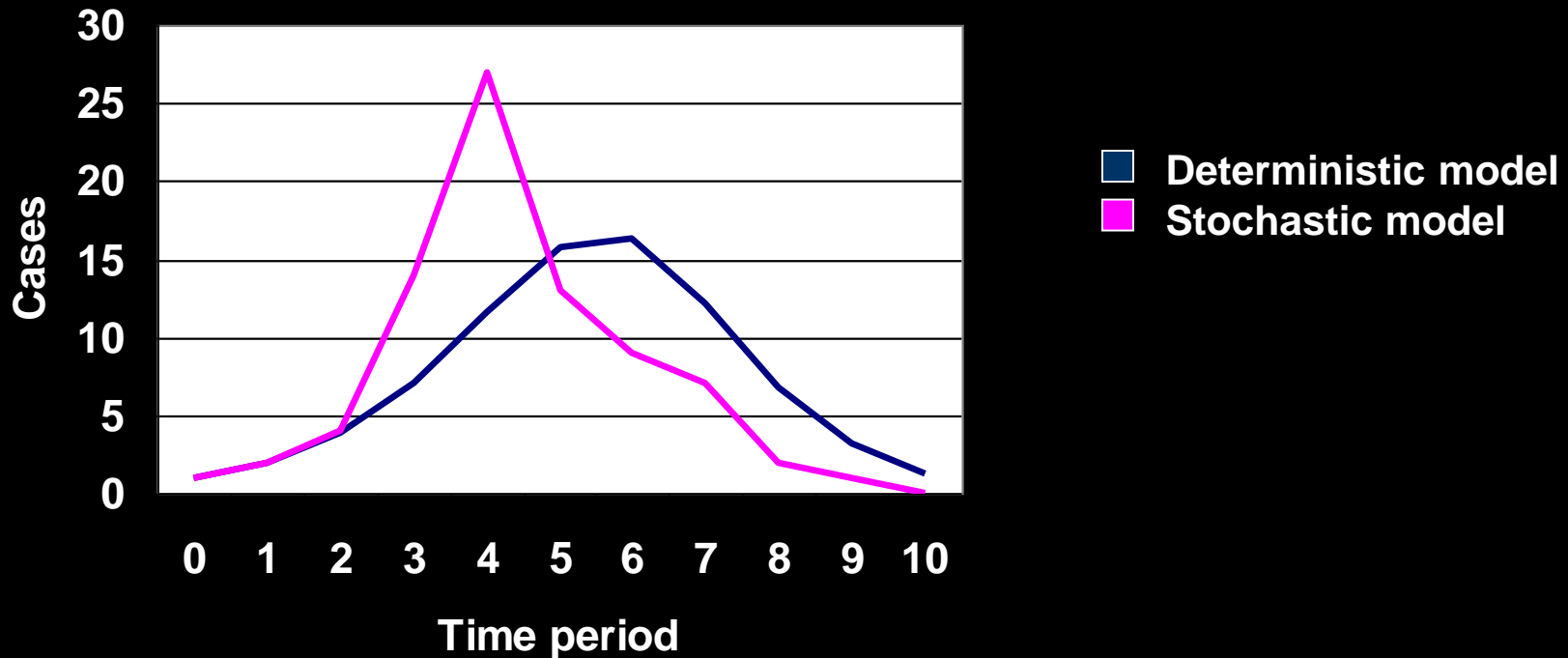
■ What is the “hidden assumption” in this model?

Model parameters:

- Population size: $N = 101$
- Initial number of cases: $C_0 = 1$
- Initial number of susceptibles: $S_0 = 100$
- $k \in \text{Binomial}(4, 0.5)$

Results of a Reed-Frost model where k is not constant (II)

Cases per time period



Model parameters:

■ Population size: $N = 101$

■ Initial number of susceptibles: $S_0 = 100$

■ Initial number of cases: $C_0 = 1$

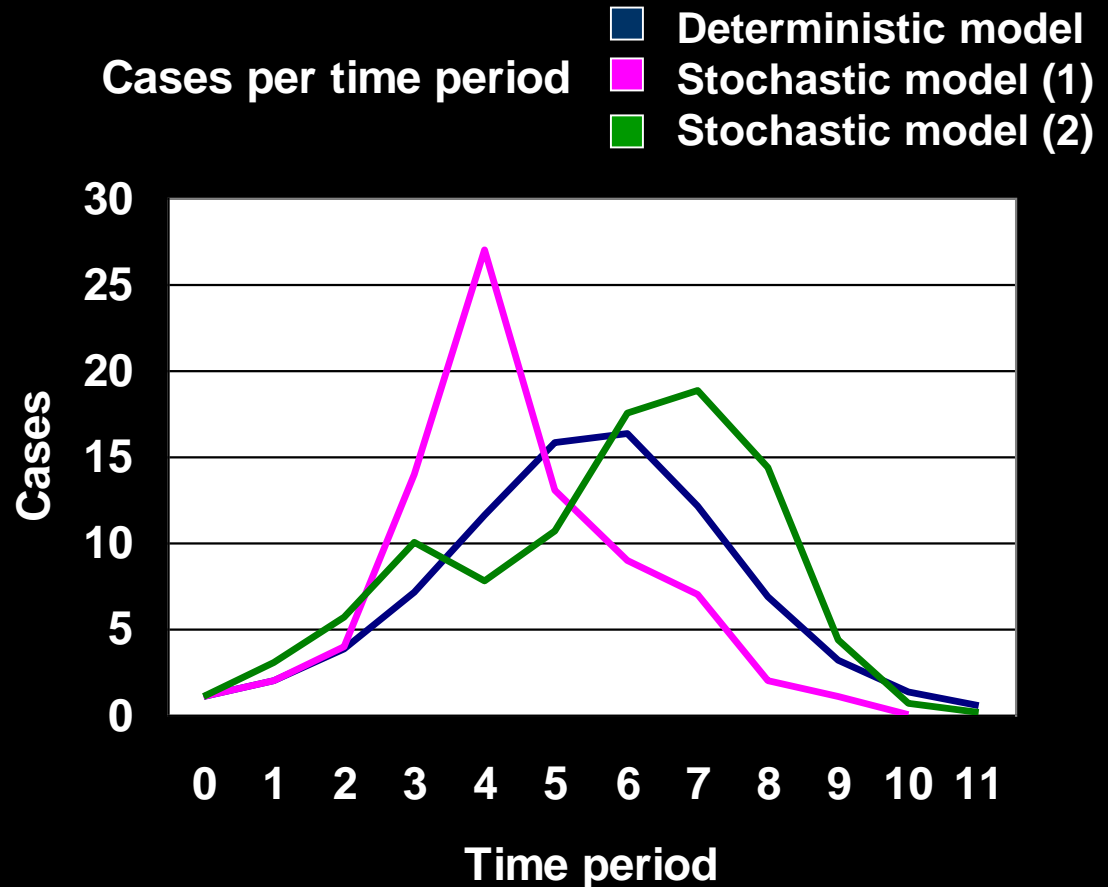
■ $k \in \text{Binomial}(4, 0.5)$

What does the stochastic model tell us so far?

- Not much
- So let's do it again, and see what happens:
- Repeat the same model, using a different sequence of values for k (which still varies according to the same binomial distribution)

Results of a Reed-Frost model where k is not constant (II)

| Time | Cases | k |
|--------------|-----------|-----|
| 0 | 1 | 3 |
| 1 | 3 | 2 |
| 2 | 6 | 2 |
| 3 | 10 | 1 |
| 4 | 8 | 2 |
| 5 | 11 | 3 |
| 6 | 17 | 3 |
| 7 | 19 | 4 |
| 8 | 14 | 3 |
| 9 | 4 | 2 |
| 10 | 1 | 2 |
| 11 | 0 | 1 |
| TOTAL | 94 | |



Model parameters:

- Population size: $N = 101$
- Initial number of cases: $C_0 = 1$
- Initial number of susceptibles: $S_0 = 100$
- $k \in \text{Binomial}(4, 0.5)$

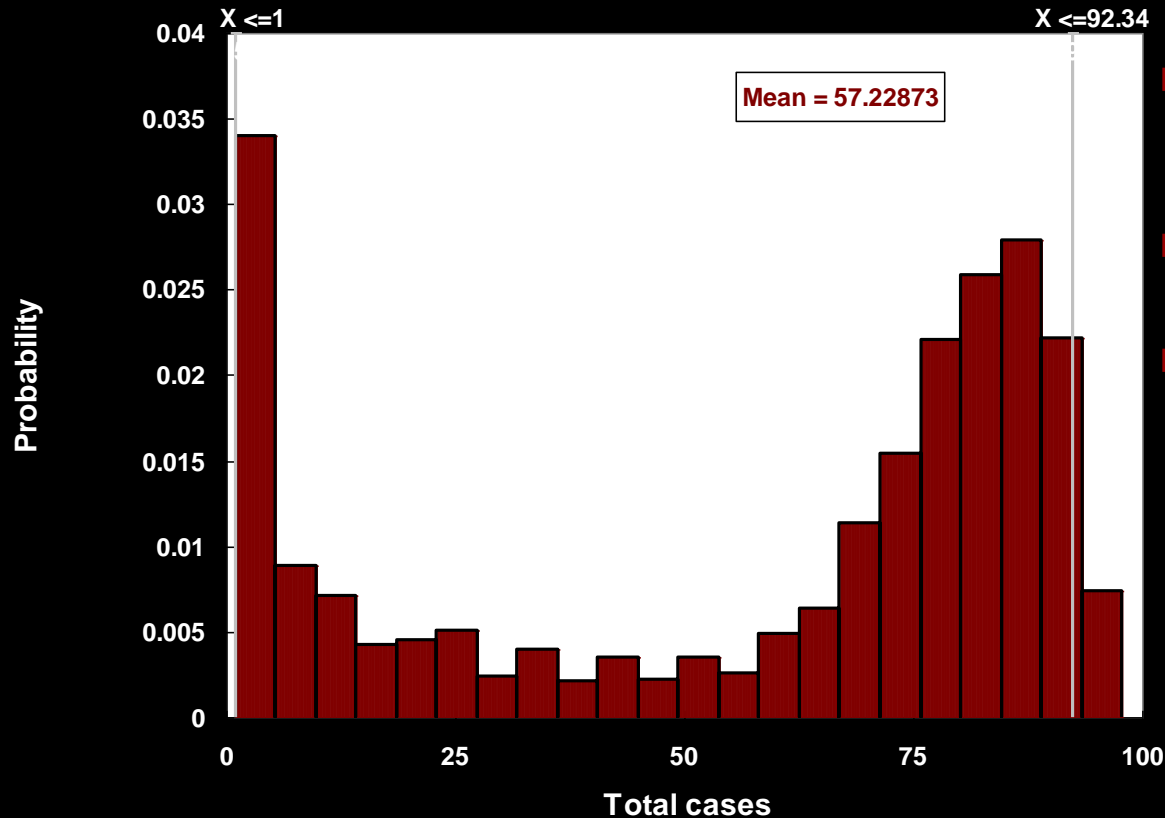
How is our stochastic model useful?

- Every time we “run” the stochastic model, we get a different result
 - This is not a bug: it’s a feature
- The result of a single “run” (also called an *iteration* or a *realization*) of a stochastic model is of little value
- When results of many iterations are considered in aggregate, we learn more than we would from any single run (even of the deterministic model)

Our stochastic model is a *Monte Carlo* simulation

- The system (spread of disease) is described by stochastic processes and probability density functions (e.g., the binomial function)
- One or more parameters (so far, just k) is allowed to vary “randomly” by *sampling* from the specified function
- The results of the simulation are expressed as a distribution, which shows the range of possible outcomes
- Monte Carlo methods can be used in situations where it is very difficult or impossible to arrive at an analytical solution
 - Complex combinations of probabilities and distributions can be used simultaneously

Total number of cases produced by our stochastic model



- What is the range in severity of possible outbreaks?
- Why is this distribution bimodal?
- Why is the mean from the stochastic model (~57) so much lower than the mean predicated by the deterministic model (~82)?

Model parameters:

■ Population size: $N = 101$

■ Initial number of susceptibles: $S_0 = 100$

■ Initial number of cases: $C_0 = 1$

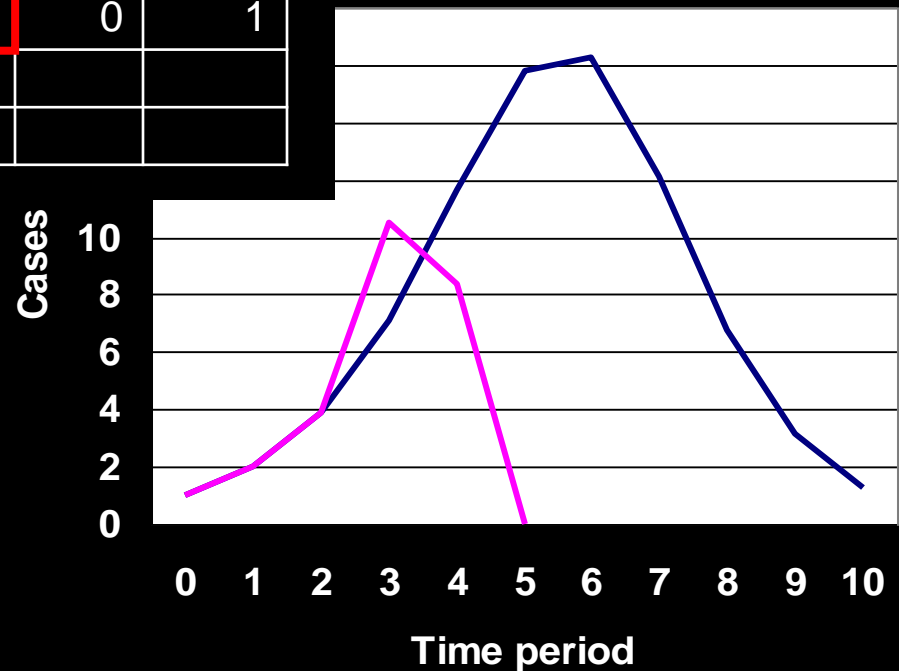
■ $k \in \text{Binomial}(4, 0.5)$

■ 10000 model iterations

Consequences of the hidden assumption

| Time | Cases | Immune | Susceptible | k | p | q |
|--------------|-----------|--------|-------------|---|------|------|
| 0 | 1 | 0 | 100 | 2 | 0.02 | 0.98 |
| 1 | 2 | 1 | 98 | 2 | 0.02 | 0.98 |
| 2 | 4 | 3 | 94 | 3 | 0.03 | 0.97 |
| 3 | 10 | 7 | 84 | 1 | 0.01 | 0.99 |
| 4 | 8 | 17 | 75 | 0 | 0 | 1 |
| 5 | 0 | 25 | 75 | | | |
| <i>Total</i> | <i>25</i> | | | | | |

■ Deterministic model
■ Stochastic model



- It is not realistic that every individual will have the same number of adequate contacts in the same time period
- It may be the best that we can do in a spreadsheet
- More detailed computer programs can be written in which the number of contacts is allowed to vary per case
 - One such model¹ produced a mean number of total cases of 78.7, much closer to the expected mean of 81.9 from the deterministic model

¹Thanks to Shaun Case

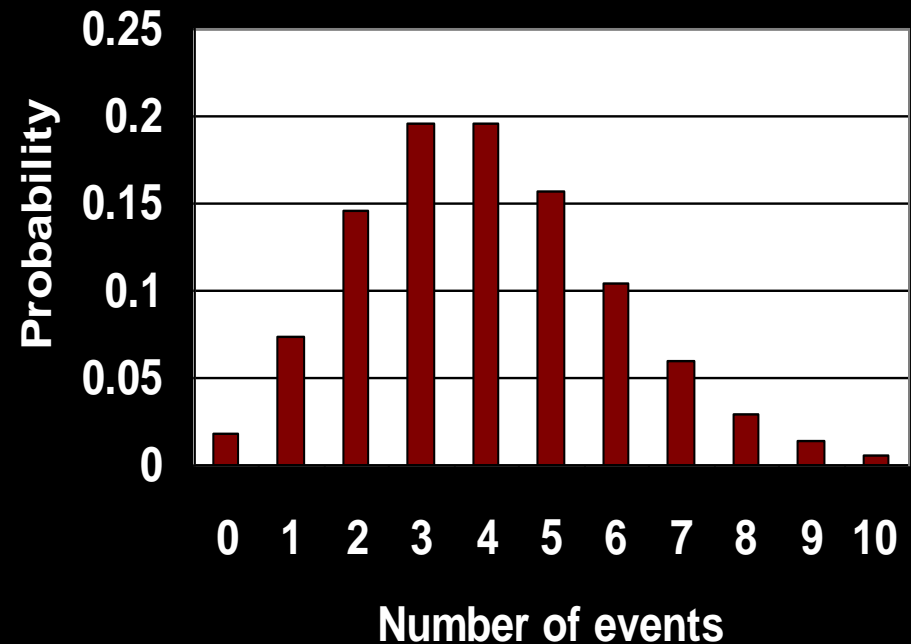
So... what else can we change?

- Recall that we said that the total number of contacts per time period (k_A) was always 4
- Could we improve the realism of our model by allowing k_A to vary?
 - If yes, how should we let it vary?

The Poisson distribution

- *Poisson distributions* are used to model the number of events that occur within a given period of time
- A Poisson distribution is specified by giving its mean

Probabilities of n events in a given time period, following a Poisson distribution with a mean of 4



Characteristics of Poisson processes

- The probability of an event (a contact) occurring over some time interval is constant
- This probability is independent of the number of events (contacts) that might have occurred in the past
- Duration between two events (contacts) is independent of the duration between any other two events

Stochastic models in which k_A varies

- One option: let k_A vary according to a Poisson distribution, and leave h fixed at 0.5:
 - $k = \text{Poisson}(4) \times 0.5$
- A second option: let k vary based on both the Poisson and Binomial distributions:
 - $k = \text{Binomial}(\text{Poisson}(4), 0.5)$
- Try them on your own, and see what happens...

Assumptions about time in a Reed-Frost model

- All models we've used so far proceed in time steps equal in duration to the average serial interval or incubation period
- Yesterday, we modeled a measles outbreak using a 12-day incubation period for the disease
- The incubation period for every individual with measles is always 12 days
- Could we allow the incubation period to vary?
 - Yes (although not in a spreadsheet model), if we know (or have an idea about) the distribution of measles incubation periods

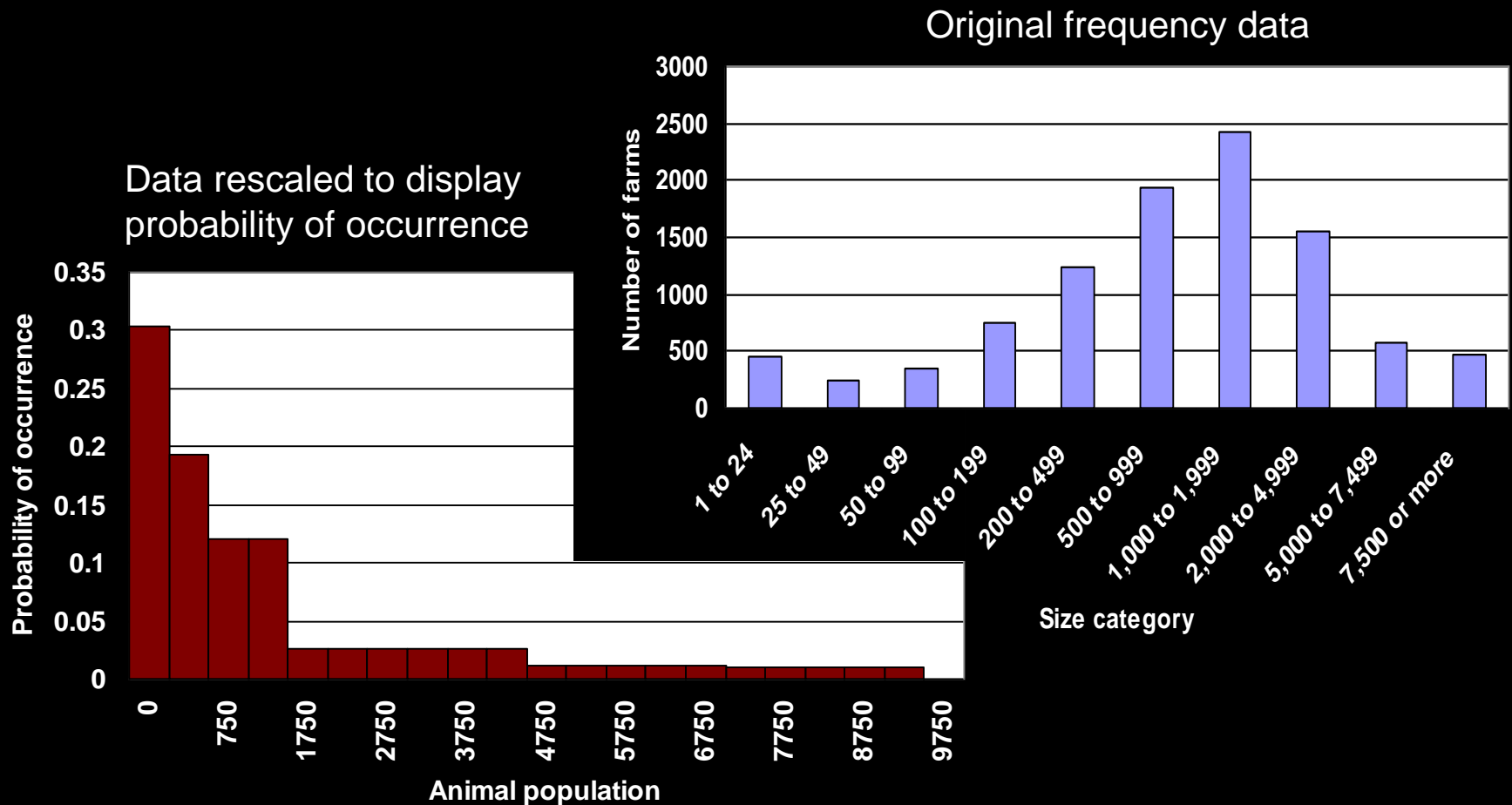
Dealing with distributions

- We said that systems in Monte Carlo models are described by *probability density* (or *probability mass*) functions
- We've now seen two such functions: the Binomial and the Poisson
- Dr. Hill will talk more about producing and using probability density functions later today
- To wrap up this session, we'll look at just a few useful probability density functions

Distributions and probability density functions

- Any kind of frequency data can be expressed as a probability density function
 - Data may be in discrete units
 - Examples: number of heads in five coin tosses; days that an infected individual is incubating disease
 - Discrete distributions have probability mass functions
 - Data may be in continuous units
 - Examples: height of four-year-old boys; milk production by dairy cows
 - Continuous distributions have probability density functions

An example: Populations of swine operations in Iowa¹



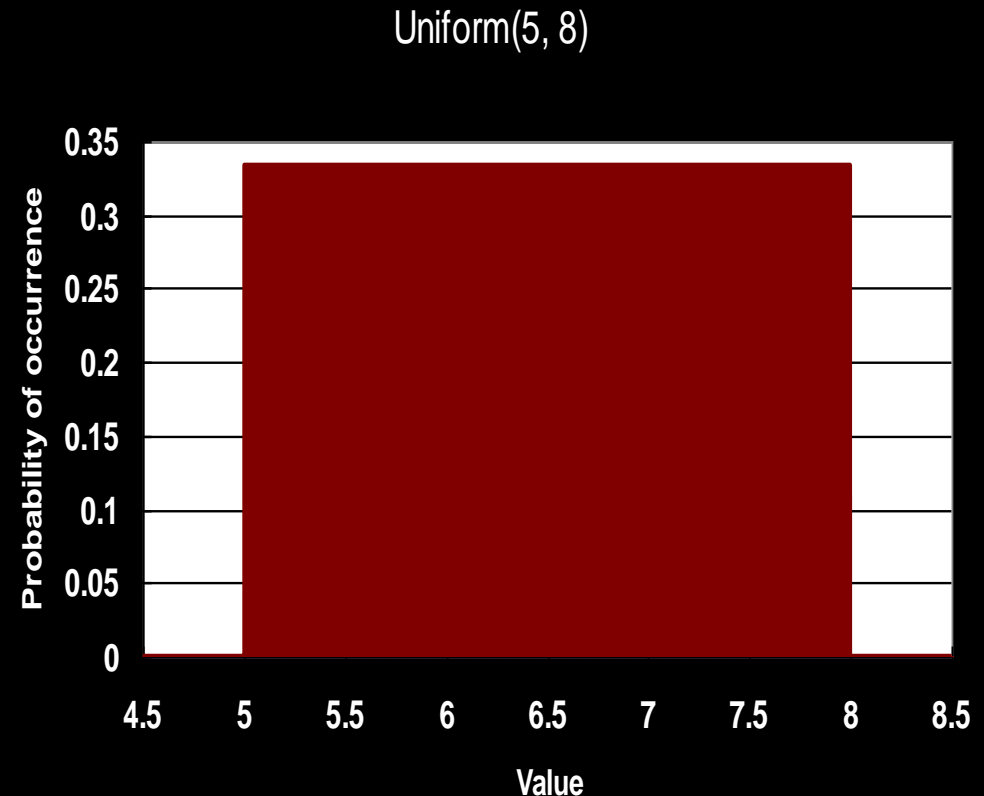
¹Original data from USDA National Agricultural Statistics Service
2002 Census of Agriculture

Several useful probability density functions

- Uniform
- Triangular
- Normal
- Exponential
- Additional probability density functions are described by Hill and Reeves (2006), Vose (2000), and in the @RISK documentation

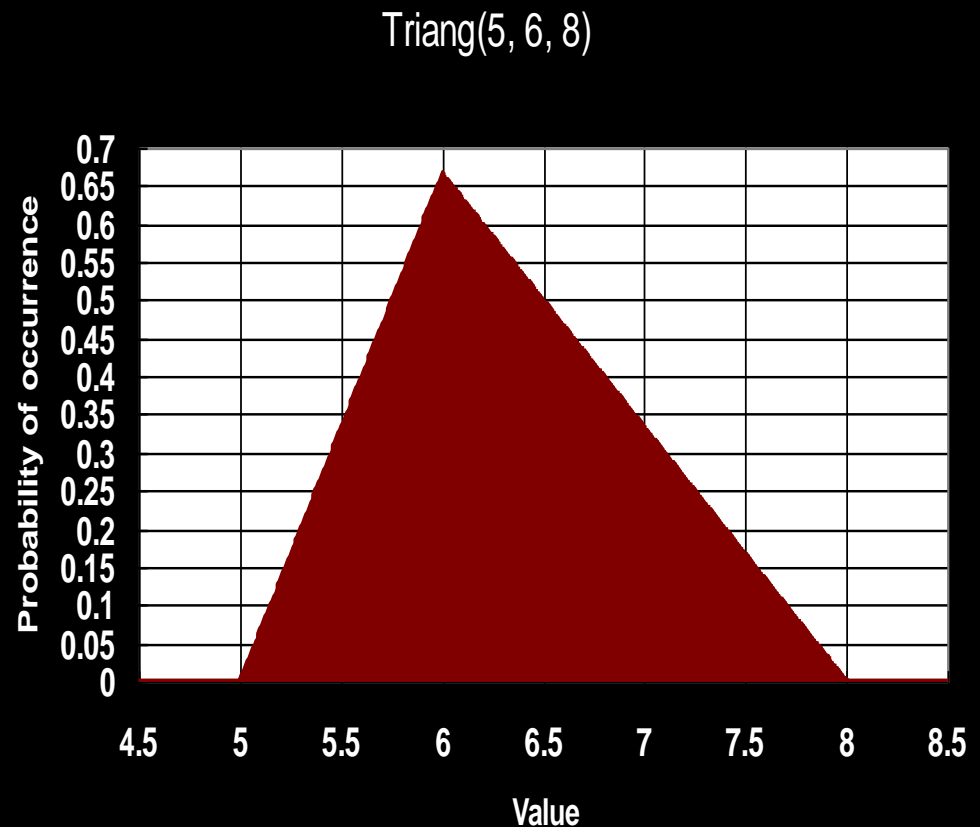
The Uniform distribution

- This distribution is defined by its minimum and maximum value
- During sampling, all values between this minimum and maximum (inclusive) have an equal probability of being selected
 - The probability of selecting a value less than the minimum or greater than the maximum is 0
- In models, Uniform distributions are generally used when very little information is available about a process or parameter



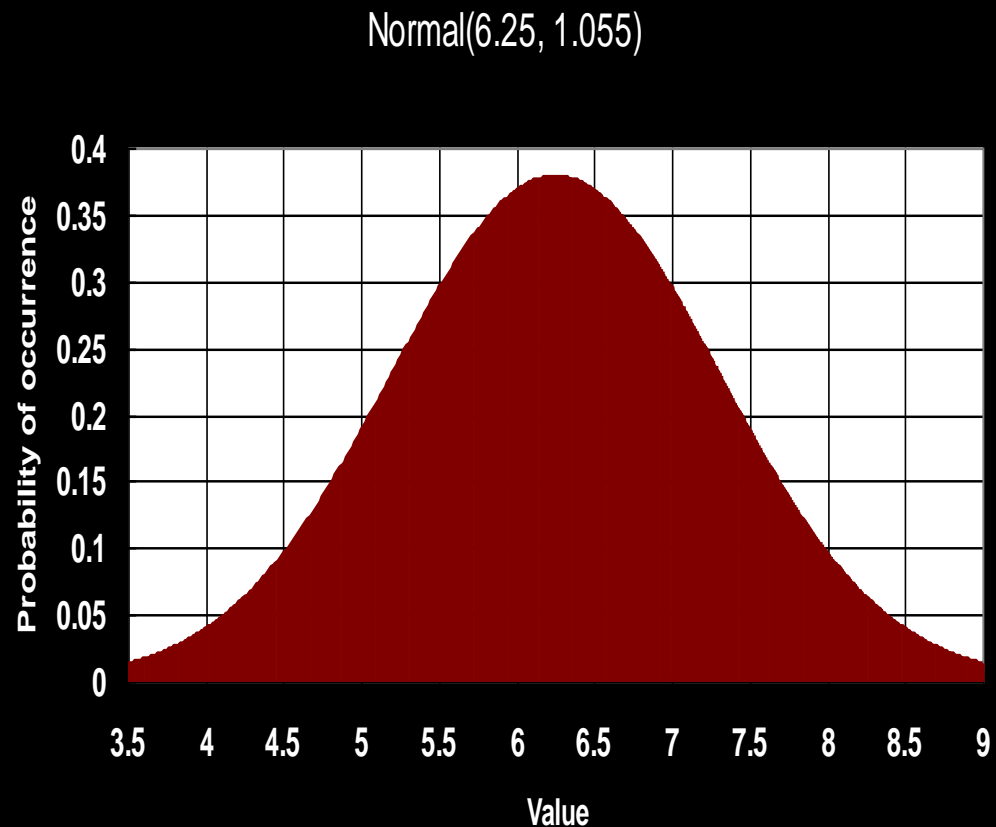
The Triangular distribution

- This distribution is defined by its minimum, most likely, and maximum value
- The distribution may be symmetrical or skewed to either the right or left
- During sampling, the most likely value has the highest probability of being selected
- Triangular distributions are used for rough modeling, where limited data is available, or where expert opinion is used to define distributions



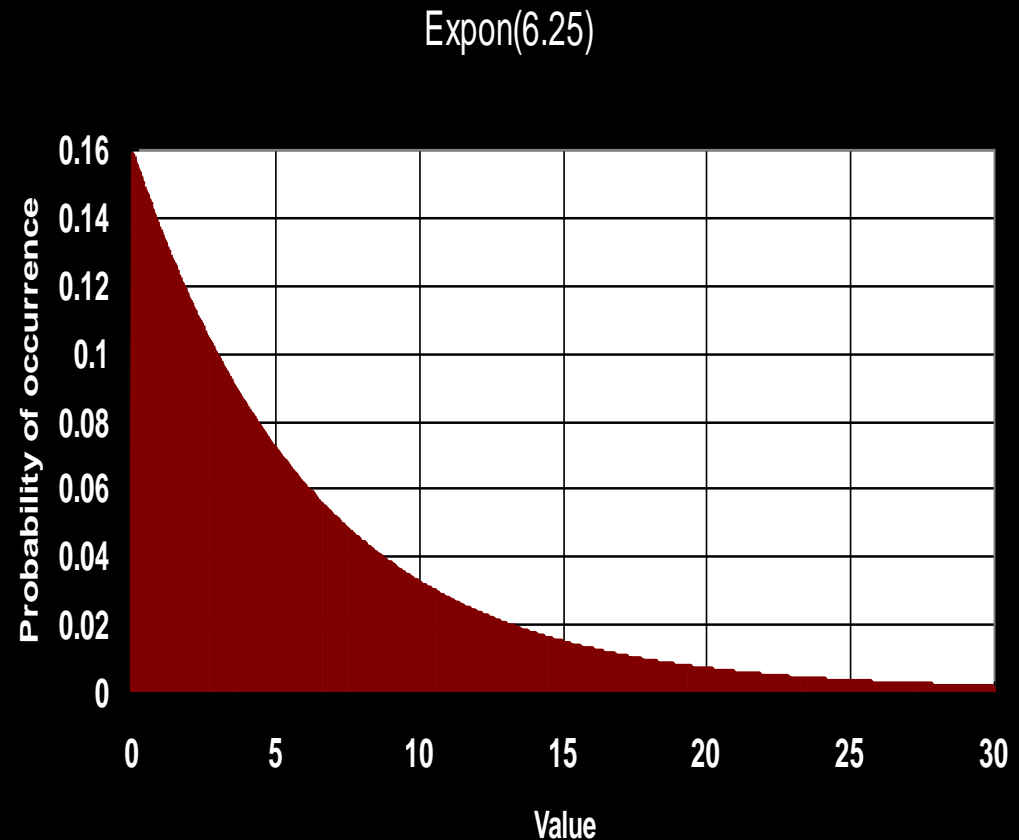
The Normal (Gaussian) distribution

- This distribution is defined by its mean and standard deviation
- Normal distributions are always symmetrical about their means
- Normal distributions are theoretically infinite in extent
- Many naturally occurring variables follow a Normal distribution
- Under the right conditions, many other distributions can be approximated with a Normal distribution



The Exponential distribution

- An Exponential distribution is defined by its mean
- Normal distributions are markedly asymmetrical
- Many differential calculus models rely on an exponential “rate of decay”
- If an event occurs with a constant probability per time unit, the time between consecutive events can be modeled with an Exponential distribution



Summary

- The Reed-Frost spreadsheet model can be modified to incorporate stochasticity
- Stochastic (Monte Carlo) models produce not just one result, but a distribution of possible results
- This distribution can be more informative than, for example, a single predicted value from a deterministic model
- Monte Carlo models are driven by probability density functions, which are used to represent variability and/or uncertainty about components of the system being modeled

Coming up...

- While the Reed-Frost model is a useful place to begin to introduce Monte Carlo methods, the degree of stochasticity that can be incorporated is limited, especially when used in a spreadsheet
- In the next session, we will examine a framework for more complex models that offers more opportunities for use of probabilistic methods
- This framework, the *state transition model*, will provide the basis for very detailed models, including those that we will create with the *North American Animal Disease Spread Model*

Recommended reading

- Carpenter, T.E., 1988. Stochastic epidemiologic modeling using a microcomputer spreadsheet package. *Preventive Veterinary Medicine* 5: 159–168. *(Slightly dated, but still a useful demonstration of stochastic Reed-Frost modeling in a spreadsheet)*
- Computational Science Education Project. 1995. Introduction to Monte Carlo Methods. An electronic book available at <http://www.phy.ornl.gov/csep/mc/mc.html>. *(A thorough introduction to the principles and algorithms used for Monte Carlo modeling)*
- Vose, D. 2000. Risk Analysis: A Quantitative Guide, 2nd ed. New York: John Wiley & Sons. *(Includes a very useful chapter on probability density functions and their applications)*

References cited

- Abbey, H., 1952. An examination of the Reed-Frost theory of epidemics. *Human Biology* 24: 201–233.
- Hill, A., Reeves, A. 2006. User's Guide for the *North American Animal Disease Spread Model*, 2nd ed. Fort Collins, Colorado: Animal Population Health Institute, Colorado State University. Available at <http://www.naadsm.org>
- Vose, D. 2000. *Risk Analysis: A Quantitative Guide*, 2nd ed. New York: John Wiley & Sons. Vose's book includes a very useful chapter on probability density functions and their applications.