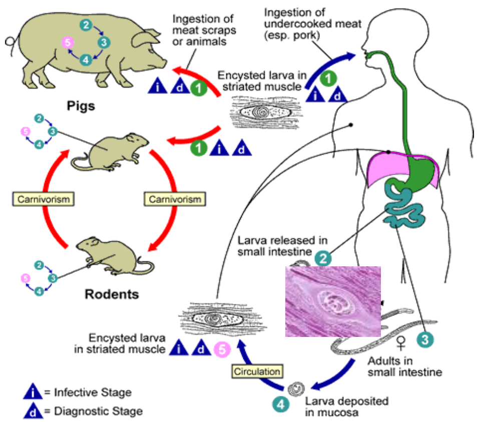
Introduction to Bayesian

(with WinBugs)

María J. Vilar
Dep. of Food Hygiene and Environmental Health



Lecture
27-31. 05. 2013

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

At the end of the course, the student should be able to:

- Differentiate and calculate conditional and unconditional probabilities
- Describe the Bayes's Theorem
- Identify the main stochastic process (CLT, Binomial, Poisson, Hypergeometric)
- Use WinBugs

Probability

- Unconditional probability   P (A) = number of individuals with A / N

P(women) = 24 / 44 , P(person with green eyes) = 15/44, P(women with blue eyes) = 6/44=0.136

	Eye colour			
	Blue	Green	Black	Total
Men	5	7	8	20
Women	6	8	10	24
Total	11	15	18	44

- Conditional probability   P(A\B) = number of individual with A / N

P(person with green eyes \women)=8/44=0.333

Probability

Independence $P(A\backslash B) = P(A)$ or if $P(B\backslash A) = P(B)$

	Gold standard test		
New test risk	Disease	Non disease	Total
Low	10	50	60
Moderate	6	30	36
High	4	20	24
Total	20	100	120

1. The probability that a individual has disease X given has a low risk is:

   P(disease X | Low Risk) = 10/60 = 0.167.

2. The probability that a individual has disease X given has a moderate risk is:

   P(disease X | Moderate Risk) = 6/36 = 0.167.

3. The probability that a individual has disease X given has a high risk is:

   P(disease X| High Risk) = 4/24 = 0.167.

Bayes' Theorem

$$P(A\backslash \mathrm{B}) = \frac{P(B\backslash \mathrm{A})P(A)}{P(B)}$$

Example

Which is the probability of disease (A), given that the patient has a positive test (B), i.e $P(A\backslash B)$?

	Diseased	Not diseased	Total
Test +	99	99	198
Test -	1	9801	9802
Total	100	9900	10000

P(A)= probability of disease (prevalence) = 1%

P(B)= probability of test positive (calculated from table)

P(B\A)= probability of test positive being disease = 99%

Bayes' Theorem

$$P(A\backslash \mathrm{B}) = \frac{P(B\backslash \mathrm{A})P(A)}{P(B)}$$

$$P(disease\backslash \text{test}+) = \frac{P(test+\backslash \text{disease}) * P(disease)}{P(test+)}$$

Example

Which is the probability of disease (A), given that the patient has a positive test (B), i.e P(A\B) ?

	Diseased	Not diseased	Total
Test +	99	99	198
Test -	1	9801	9802
Total	100	9900	10000

P(A)= probability of disease (prevalence) = 1%

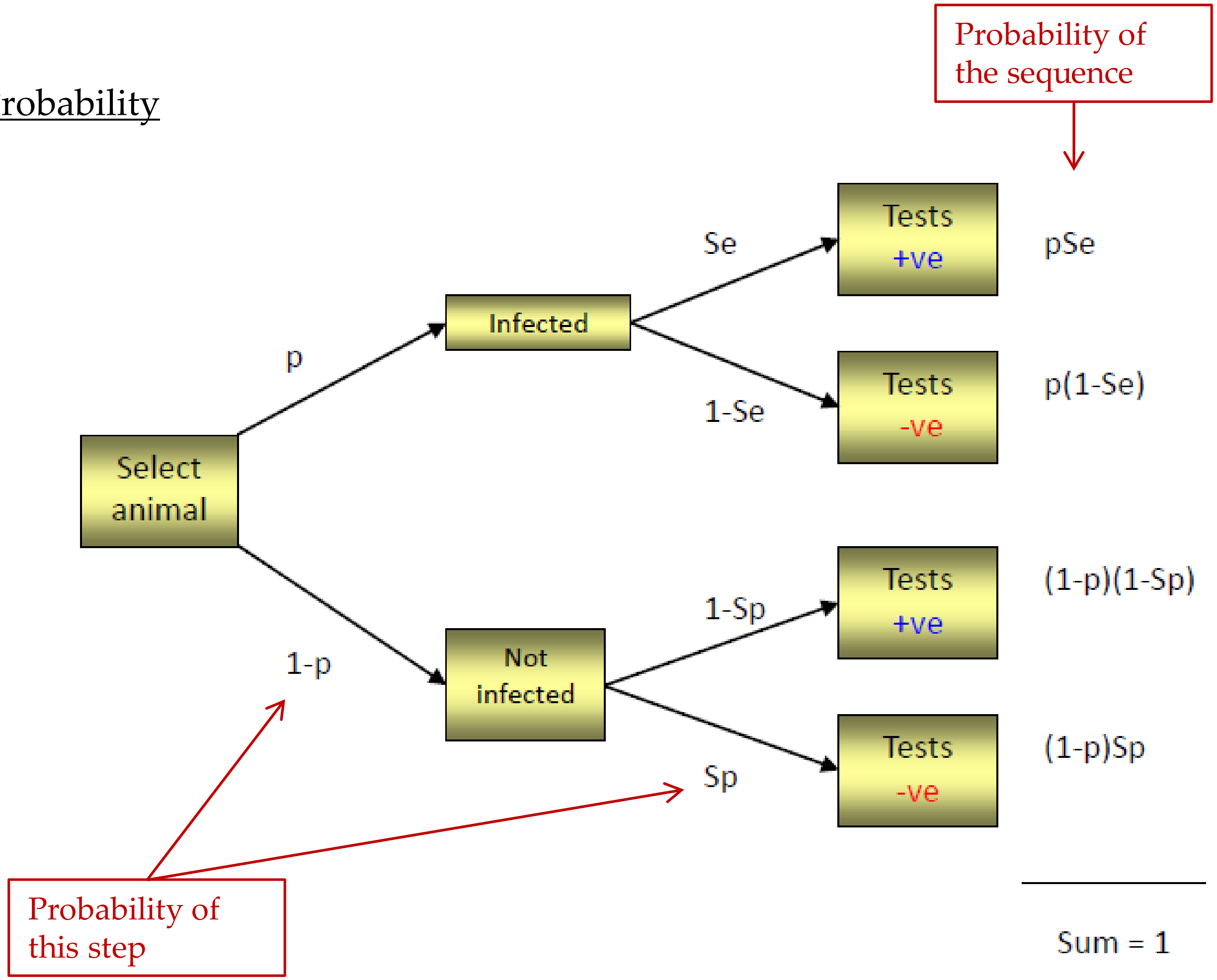
P(B)= probability of test positive (calculated from table)

P(B\A)= probability of test positive being disease = 99%

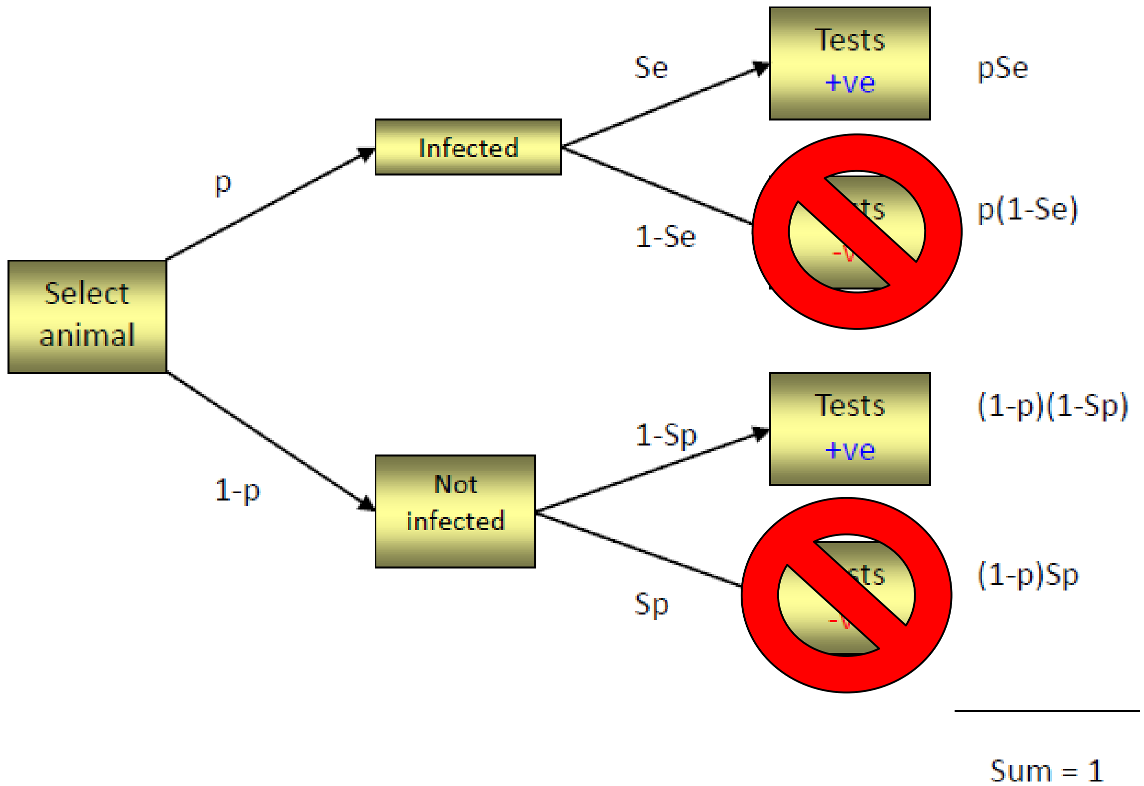
P(B)= 198/10000= 0.0198

P(A\B) = (0.99/0.01) / 0.0198 = 0.5 or 50%

Probability



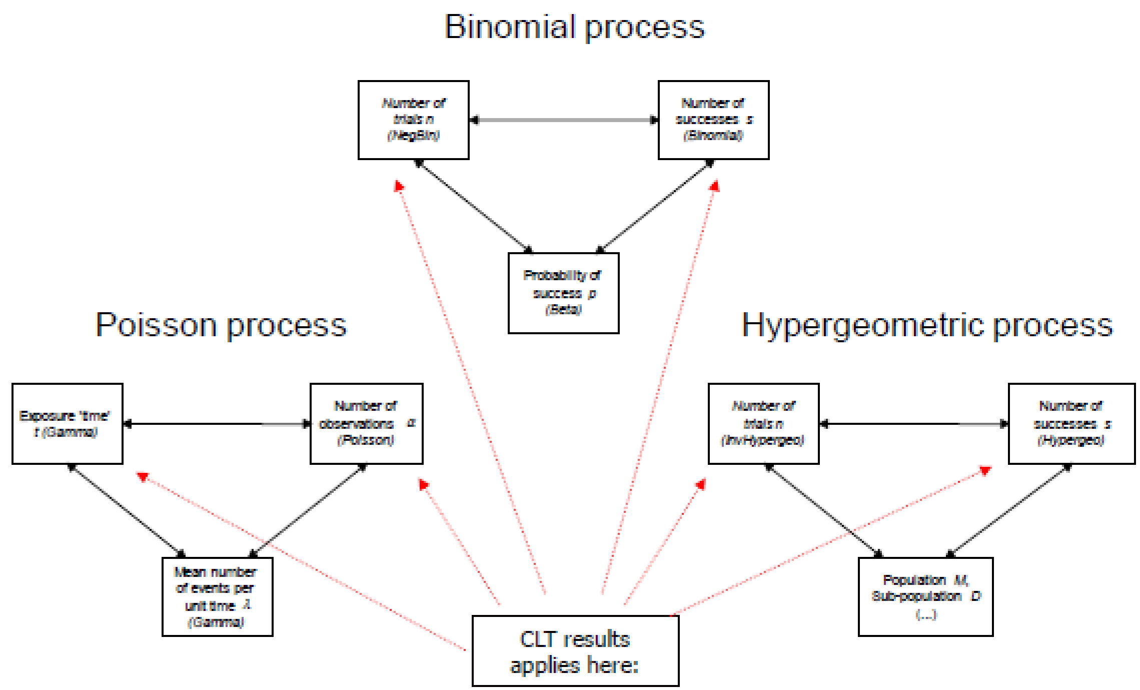
Probability



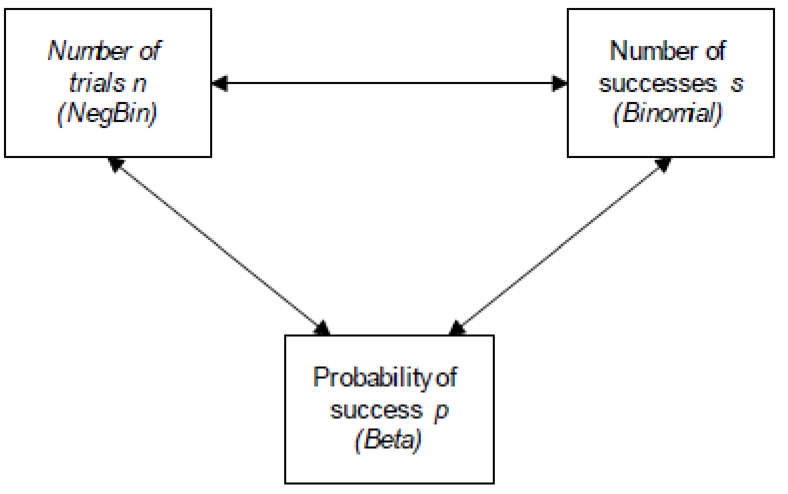
$$P(\text{infected} \mid \text{+ve test}) = \frac{pSe}{pSe + (1-p)(1-Sp)}$$

Stochastic processes

- Central Limit Theorem (CLT)
- Binomial
- Poisson
- Hypergeometric



Binomial Process



- **n** independent trials
- each trial has the same probability **p**
- .. resulting in **s** successes

Distributions

s = binomial (n,p)

n= s+ negbinomial (s,p)

p= beta (s+1, n-s+1)

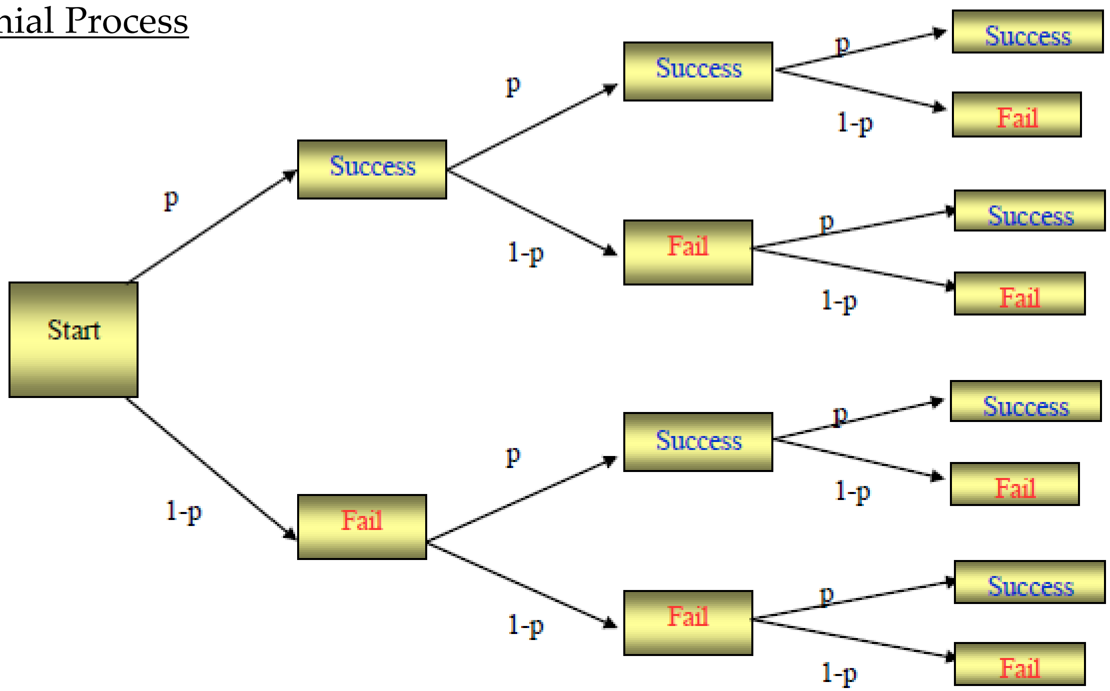
Example: Test for contaminated burger

Trial: burger

Prevalence: probability of contamination

Success: contaminated burger

Binomial Process



Binomial equation

$$P(s,n,p) = \binom{n}{s} p^s (1-p)^{n-s}$$

Binomial probability mass function

$$P(s) = \frac{n!}{s!\,(n-s)!} p^s (1-p)^{n-s}$$

Example

Imagine that 80% of burgers sold in restaurants from company X are contaminated with E.coli. If we analyse 10 burgers, what is the probability of find exactly seven contaminated burgers?

$$P(s) = \frac{n!}{s!\,(n-s)!} p^s (1-p)^{n-s}$$

We know that:

10 analysis (n) are performed,

7 out of 10 analysis are sucesful (s),

probability of sucess for each analysis is 0.8 (p).

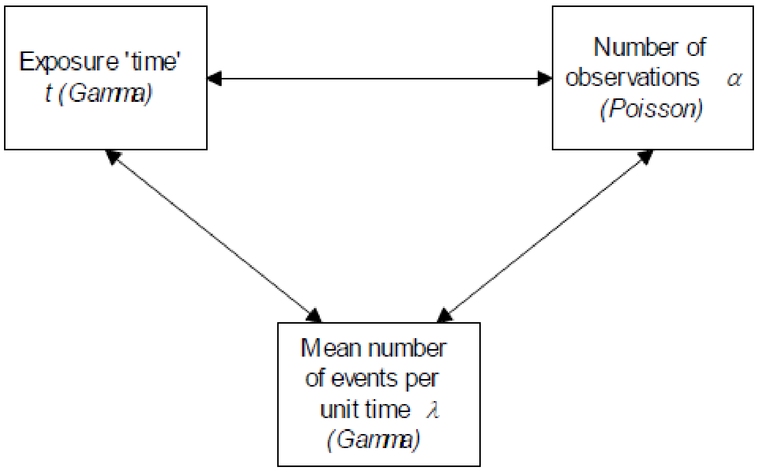
$$P(7 contaminated\ burger) = \frac{10!}{7!\,(10-7)!} 0.8^7 (1-0.8)^{10-7} = 0.201$$

Quiz

What are the three assumptions behing a binomial process?

a. All trials must occur together
b. All trials are independent
c. All trials have the same probability
d. There must be at least one success
e. Every trial must be either a success or a failure
f. Not all trials can be successes

Poisson Process



- opportunity **l** of events to occur
- likelihood per unit exposure **λ**
- .. resulting in **α** events occurring in **t**

α = Poisson (λ*t) , if P (α=0) = Exp(-λt)

t = Gamma (α , β) , where $\beta = 1/\lambda$

λ = Gamma (α / 1/t)

Example:

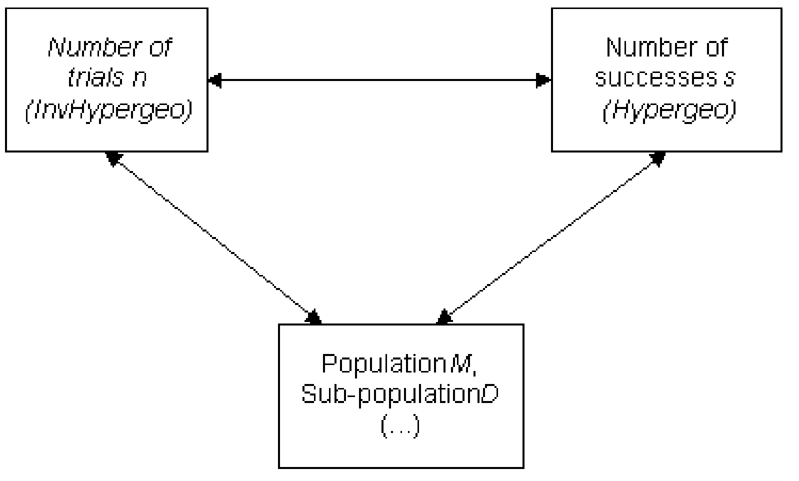
Giardia cysts consumed in litters of water

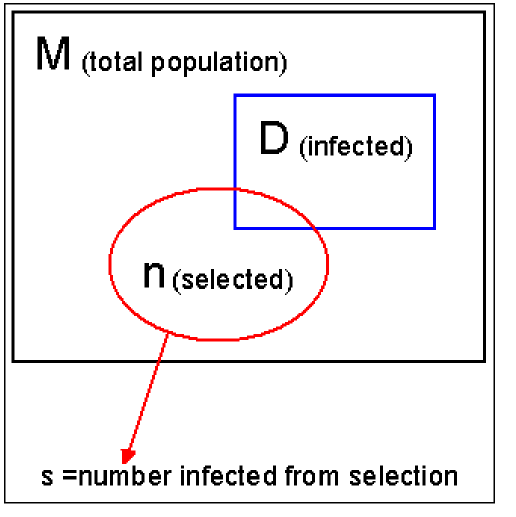
Outbreaks in a year

Poisson <--> Binomial

$n \to \infty, p \to 0, np \to \lambda t$

<u>Hypergeometric Process</u>

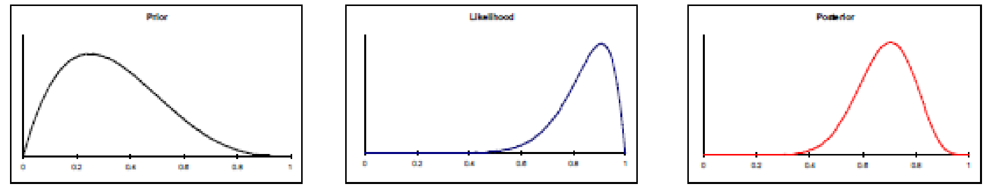




Example:

Herd testing for pathogen X

Stochastic analyses



What we knew before + What the data tells us = What we know now

$$f(\theta|X) = \frac{\pi(\theta).l(X|\theta)}{\int \pi(\theta).l(X|\theta).d\theta} \quad \Longrightarrow \quad f(\theta|X) \propto \pi(\theta).l(X|\theta)$$

Prior distirbution Π (θ) P(A)

Likelihood L (X|θ) P(B\A)

the likelihood of the unknown model parameters given the data

conditional probability of observations B, given A

Posterior F (θ|X) P(A\B)

posterior probability distribution,

the probability of observing an outcome A given the known parameter B

References

Spiegelhalter, D. et al. WinBugs User Manual. 2003.

Ntzoufras, I. Bayesian Modeling using WinBugs. Wiley, 2009

The WinBugs project website.

http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml

http://www.epixanalytics.com/modelassist/AtRisk/Model_Assist.htm#Introduction.htm