Probability Distributions

Discrete Data Modeling

Modeling Randomness-I

- What is the probability that you flip exactly 4 heads in 10 coin tosses assuming a coin having following probability distributions:
    a. p(H)=0.5 & p(T)=0.5
    b. p(H)=0.3 & p(T)=0.7

Binomial Distribution

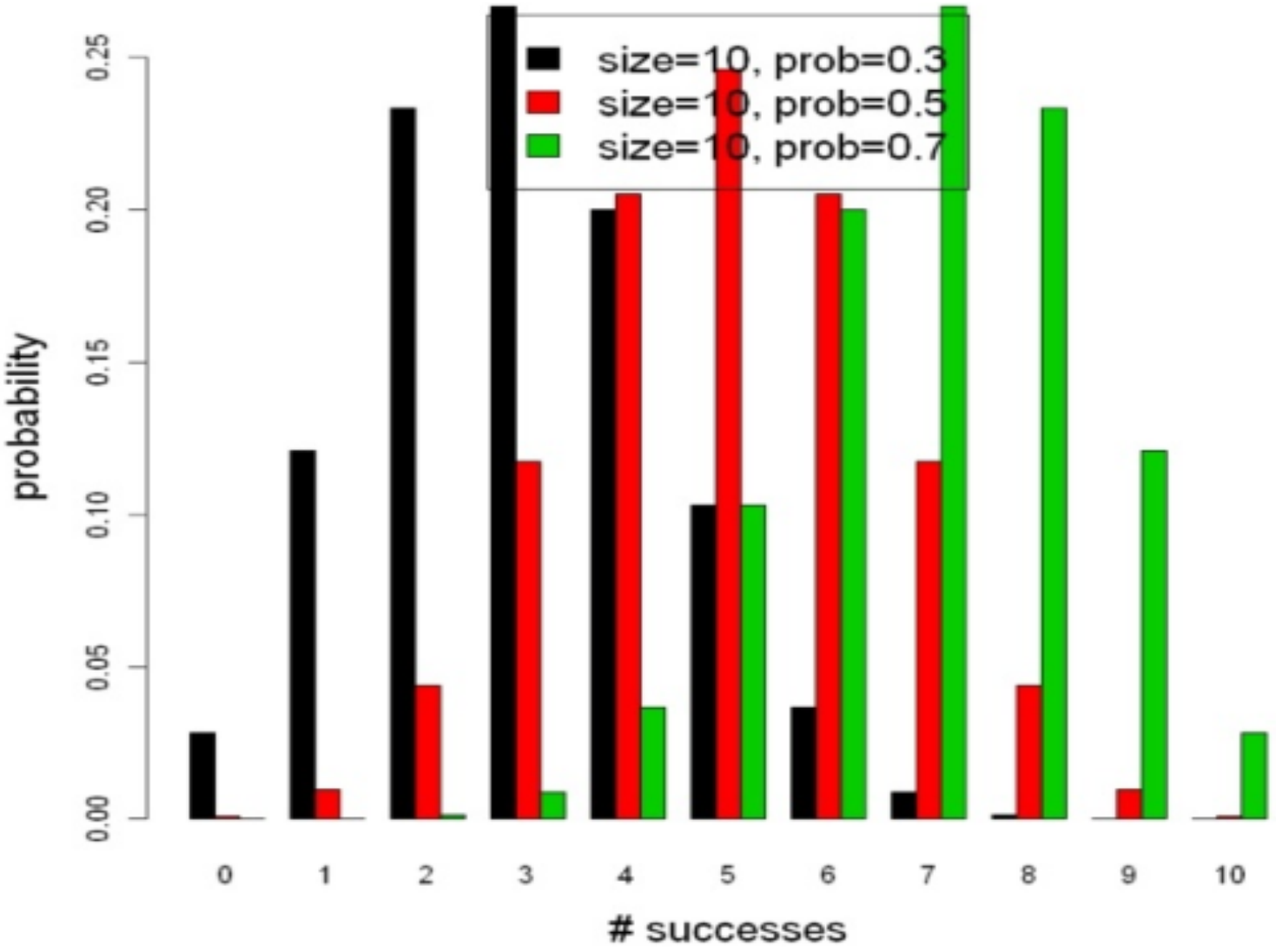
- It gives us the number of successes(x) in a sample, with fixed number of trials(n) and equal probability of success(p) in every trial.

$$\binom{N}{x} p^x (1-p)^{N-x}$$

N = trial size
p = per trial prob(success)
x = #successes

- Used when x has an upper limit. When n is large and p is small, it approaches poisson.

Binomial Distribution



Modeling Randomness-I

- Assuming a coin having following probability distribution: p(H)=0.7 & p(T)=0.3, Find the probability that you flip
    a. atmost 4 heads in 10 coin tosses?
    b. atleast 4 heads in 10 coin tosses?

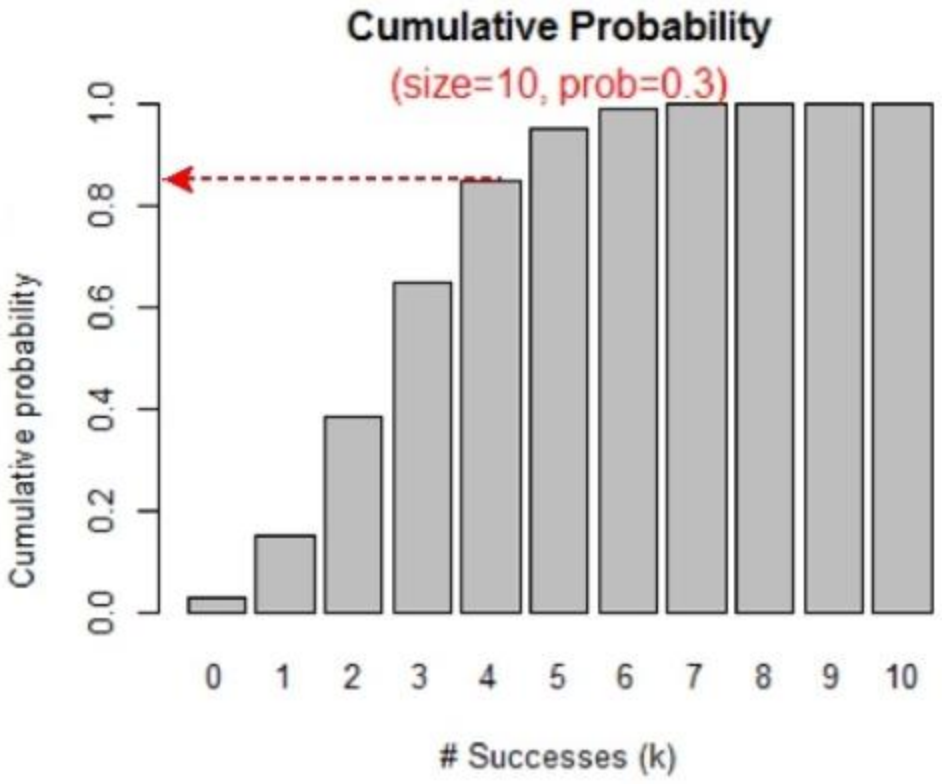
Solution

- P(X <= 4) = B(0, 10, 0.3) + B(1, 10, 0.3) + B(2, 10, 0.3) + B(3, 10, 0.3) + B(4, 10, 0.3)

Cumulative probability

$$f(x) = \text{Prob}(X \leq x)$$

- Denotes probability of x being less than or equal to any particular value (basis for *p*-values)



pbinom(x=4,size=10,prob=0.3)
= 0.85

Solution

- Let X be a random variable that represents the number of heads tossed in 5 tosses.
- $P(X >= 4) = 1 - P(X <= 3)$
- **In R:** 1 - pbinom(3,10,0.3)

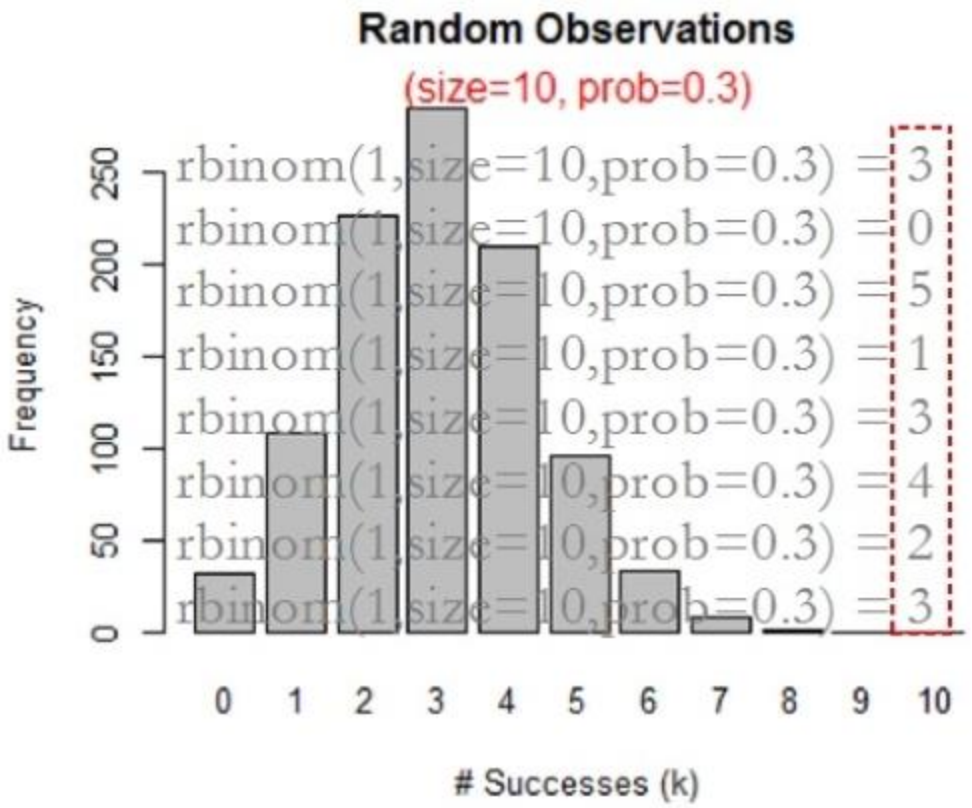
Modeling Randomness-I

- Generate random sample from binomial distribution.

Solution

Random numbers:

- Random draws from the specified distribution
- Each draw will produce a different result, which gives rise to the probability distribution
- Extremely useful for simulations



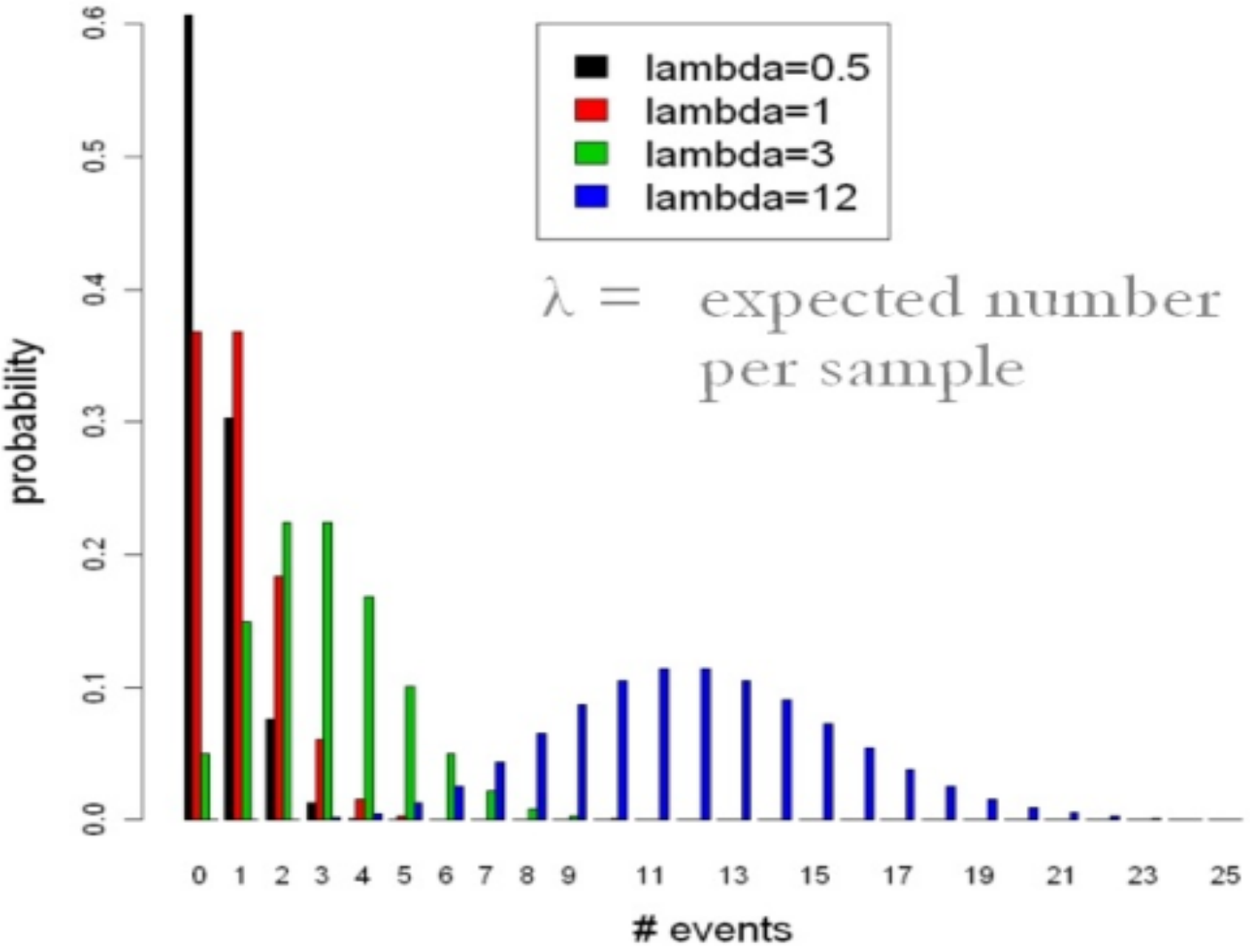
Poisson Distribution Pattern

Poisson Pattern

- It gives us the number of events in a given unit of sampling effort if each event is independent.

- Used when you expect the number of events to be effectively unlimited.

Poisson Pattern



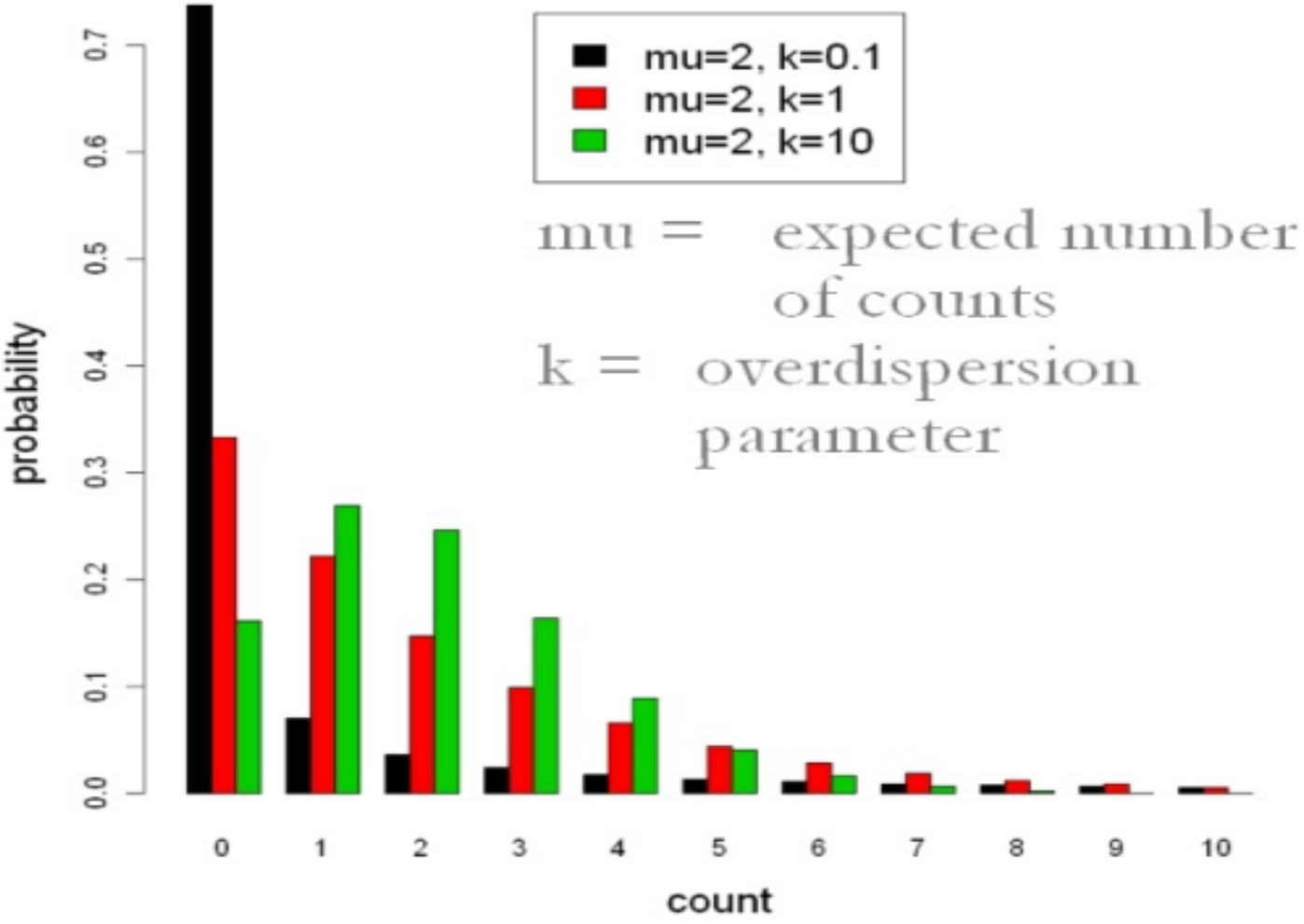
Negative Binomial Distribution Pattern

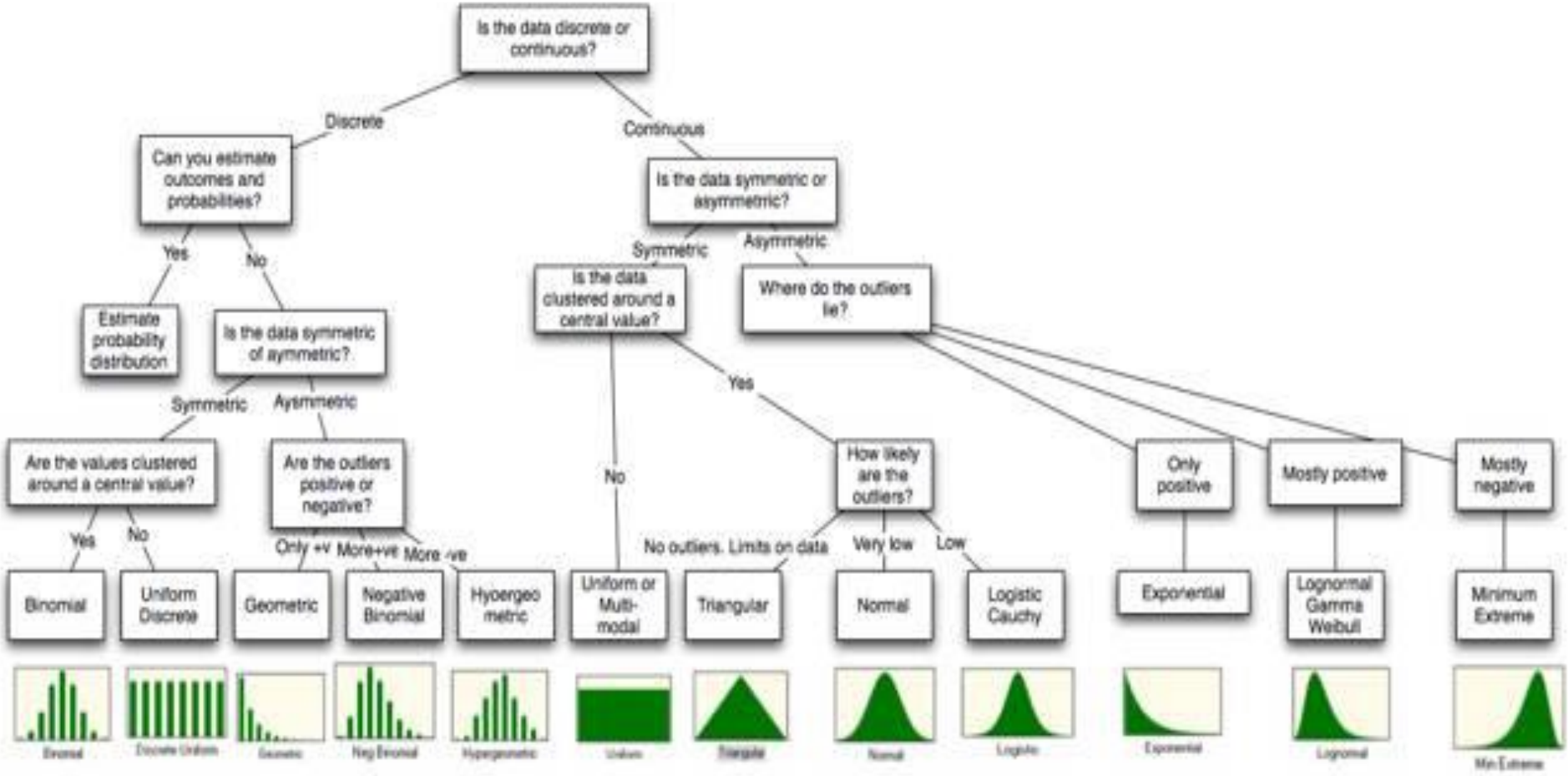
Negative Binomial Pattern

- It gives us the number of failures before a predetermined number of successes(n) occurs in a sample with equal probability of success in every trial(prob).

- Used as with poisson but with heterogeneity in data(overdispersion: variance > mean).

Negative Binomial Pattern




Is the data discrete or continuous?
Discrete
Continuous
Can you estimate outcomes and probabilities?
Is the data symmetric or asymmetric?
Yes
No
Symmetric
Asymmetric
Estimate probability distribution
Is the data symmetric of aymmetric?
Is the data clustered around a central value?
Where do the outliers lie?
Symmetric
Aytmmetric
Yes
Are the values clustered around a central value?
Are the outliers positive or negative?
No
How likely are the outliers?
Only positive
Mostly positive
Mostly negative
Yes
No
Only +v
More+ve
More -ve
No outliers. Limits on data
Very low
Low
Binomial
Uniform Discrete
Geometric
Negative Binomial
Hyoergeo metric
Uniform or Multi-modal
Triangular
Normal
Logistic Cauchy
Exponential
Lognormal Gamma Weibull
Minimum Extreme