

## Lecture 2

# Binomial and Poisson Probability Distributions

### Binomial Probability Distribution

- Consider a situation where there are only two possible outcomes (a Bernoulli trial)

- ★ Example:

- ◆ flipping a coin

- ☞ head or tail

- ◆ rolling a dice

- ☞ 6 or not 6 (i.e. 1, 2, 3, 4, 5)

- ★ Label the possible outcomes by the variable  $k$

- ☞ find the probability  $P(k)$  for event  $k$  to occur

- ★ Since  $k$  can take on only 2 values we define those values as:

- $k = 0$  or  $k = 1$

- ◆ let  $P(k = 0) = q$  (remember  $0 \leq q \leq 1$ )

- ◆ *something* must happen

- ☞  $P(k = 0) + P(k = 1) = 1$

- ☞  $P(k = 1) = p = 1 - q$

- ◆ write the probability distribution  $P(k)$  as:

- $P(k) = p^k q^{1-k}$  (Bernoulli distribution)

- ◆ coin toss: define probability for a head as  $P(1)$

- ☞  $P(1) = 0.5$

- ◆ dice rolling: define probability for a six to be rolled as  $P(1)$

- ☞  $P(1) = 1/6$

- ☞  $P(0) = 5/6$  (not a six)



James Bernoulli (Jacob I)  
born in Basel, Switzerland  
Dec. 27, 1654-Aug. 16, 1705  
He is one 8 mathematicians  
in the Bernoulli family  
(from Wikipedia).

- What is the mean ( $\mu$ ) of  $P(k)$ ?

$$\mu = \frac{\sum_{k=0}^1 kP(k)}{\sum_{k=0}^1 P(k)} = \frac{0 \cdot q + 1 \cdot p}{q + p} = p$$

- What is the Variance ( $\sigma^2$ ) of  $P(k)$ ?

$$\sigma^2 = \frac{\sum_{k=0}^1 k^2 P(k)}{\sum_{k=0}^1 P(k)} - \mu^2 = 0^2 \cdot P(0) + 1^2 \cdot P(1) - \mu^2 = p - p^2 = p(1 - p) = pq$$

- Suppose we have  $N$  trials (e.g. we flip a coin  $N$  times)

☞ what is the probability to get  $m$  successes (= heads)?

- Consider tossing a coin twice. The possible outcomes are:

★ no heads:  $P(m = 0) = q^2$

★ one head:  $P(m = 1) = qp + pq$  (toss 1 is a tail, toss 2 is a head or toss 1 is head, toss 2 is a tail)  
 $= 2pq$

★ two heads:  $P(m = 2) = p^2$

two outcomes because we don't care which of the tosses is a head

★  $P(0) + P(1) + P(2) = q^2 + 2pq + p^2 = (q + p)^2 = 1$

- We want the probability distribution function  $P(m, N, p)$  where:

$m$  = number of success (e.g. number of heads in a coin toss)

$N$  = number of trials (e.g. number of coin tosses)

$p$  = probability for a success (e.g. 0.5 for a head)

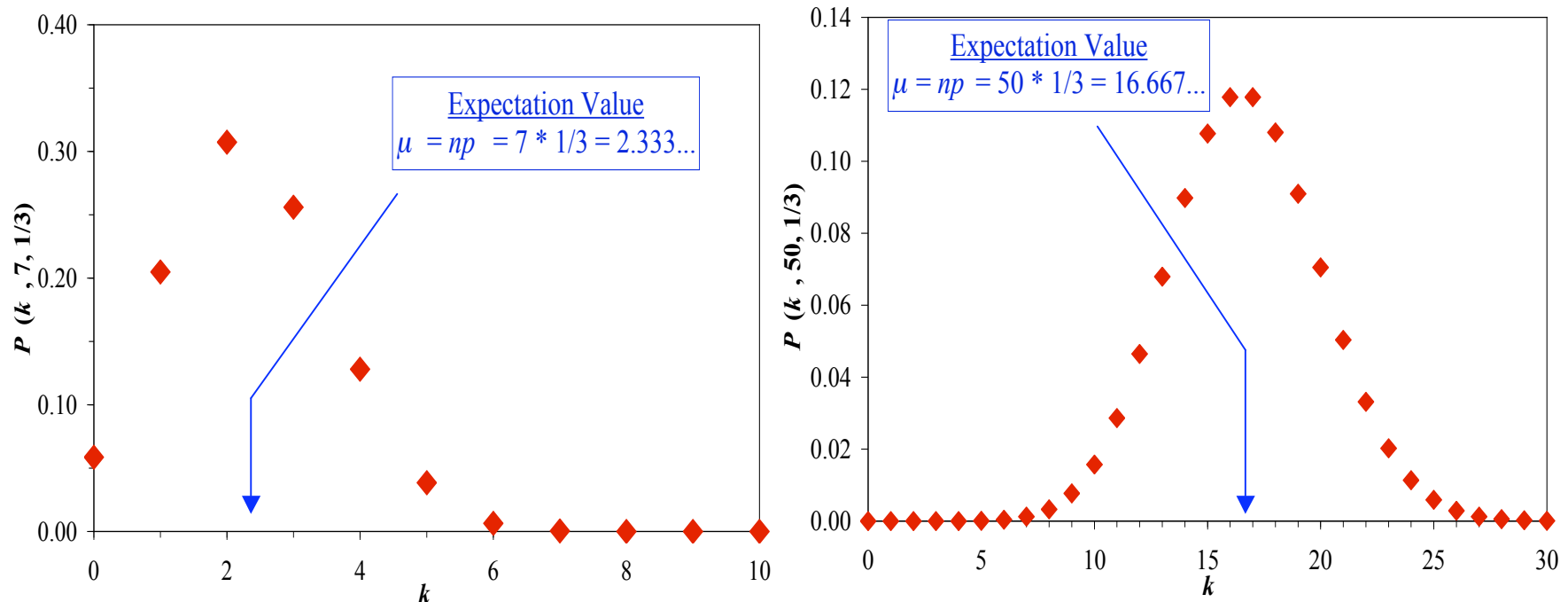
- If we look at the three choices for the coin flip example, each term is of the form:  
 $C_m p^m q^{N-m}$   $m = 0, 1, 2$ ,  $N = 2$  for our example,  $q = 1 - p$  always!
  - ★ coefficient  $C_m$  takes into account the number of ways an outcome can occur regardless of order
  - ★ for  $m = 0$  or  $2$  there is only one way for the outcome (both tosses give heads or tails):  $C_0 = C_2 = 1$
  - ★ for  $m = 1$  (one head, two tosses) there are two ways that this can occur:  $C_1 = 2$ .
- Binomial coefficients: number of ways of taking  $N$  things  $m$  at a time
 
$$C_{N,m} = \binom{N}{m} = \frac{N!}{m!(N-m)!}$$
  - ★  $0! = 1! = 1$ ,  $2! = 1 \cdot 2 = 2$ ,  $3! = 1 \cdot 2 \cdot 3 = 6$ ,  $m! = 1 \cdot 2 \cdot 3 \cdots m$
  - ★ Order of things is not important
    - ◆ e.g. 2 tosses, one head case ( $m = 1$ )
      - we don't care if toss 1 produced the head or if toss 2 produced the head
  - ★ Unordered groups such as our example are called *combinations*
  - ★ Ordered arrangements are called *permutations*
  - ★ For  $N$  distinguishable objects, if we want to group them  $m$  at a time, the number of permutations:
 
$$P_{N,m} = \frac{N!}{(N-m)!}$$
    - ◆ example: If we tossed a coin twice ( $N = 2$ ), there are two ways for getting one head ( $m = 1$ )
    - ◆ example: Suppose we have 3 balls, one white, one red, and one blue.
      - Number of possible pairs we could have, keeping track of order is 6 (rw, wr, rb, br, wb, bw):
 
$$P_{3,2} = \frac{3!}{(3-2)!} = 6$$
      - If order is **not** important (rw = wr), then the binomial formula gives
 
$$C_{3,2} = \frac{3!}{2!(3-2)!} = 3$$

number of two-color combinations

- Binomial distribution: the probability of  $m$  success out of  $N$  trials:

$$P(m, N, p) = C_{N,m} p^m q^{N-m} = \binom{N}{m} p^m q^{N-m} = \frac{N!}{m!(N-m)!} p^m q^{N-m}$$

- ♦  $p$  is probability of a success and  $q = 1 - p$  is probability of a failure
- ♦ Consider a game where the player bats 4 times:
  - ★ probability of 0/4 =  $(0.67)^4 = 20\%$
  - ★ probability of 1/4 =  $[4!/(3!1!)](0.33)^1(0.67)^3 = 40\%$
  - ★ probability of 2/4 =  $[4!/(2!2!)](0.33)^2(0.67)^2 = 29\%$
  - ★ probability of 3/4 =  $[4!/(1!3!)](0.33)^3(0.67)^1 = 10\%$
  - ★ probability of 4/4 =  $[4!/(0!4!)](0.33)^4(0.67)^0 = 1\%$
  - ★ probability of getting at least one hit =  $1 - P(0) = 0.8$



- To show that the binomial distribution is properly normalized, use Binomial Theorem:

$$(a+b)^k = \sum_{l=0}^k \binom{k}{l} a^{k-l} b^l$$

$$\sum_{m=0}^N P(m, N, p) = \sum_{m=0}^N \binom{N}{m} p^m q^{N-m} = (p+q)^N = 1$$

☞ binomial distribution is properly normalized

- Mean of binomial distribution:

$$\mu = \frac{\sum_{m=0}^N m P(m, N, p)}{\sum_{m=0}^N P(m, N, p)} = \sum_{m=0}^N m P(m, N, p) = \sum_{m=0}^N m \binom{N}{m} p^m q^{N-m}$$

- ★ A cute way of evaluating the above sum is to take the derivative:

$$\frac{\partial}{\partial p} \left[ \sum_{m=0}^N \binom{N}{m} p^m q^{N-m} \right] = 0$$

$$\sum_{m=0}^N m \binom{N}{m} p^{m-1} q^{N-m} - \sum_{m=0}^N \binom{N}{m} p^m (N-m)(1-p)^{N-m-1} = 0$$

$$p^{-1} \sum_{m=0}^N m \binom{N}{m} p^m q^{N-m} = N(1-p)^{-1} \sum_{m=0}^N \binom{N}{m} p^m (1-p)^{N-m} - (1-p)^{-1} \sum_{m=0}^N m \binom{N}{m} p^m (1-p)^{N-m}$$

$$p^{-1} \mu = N(1-p)^{-1} \cdot 1 - (1-p)^{-1} \mu$$

$$\mu = Np$$

- Variance of binomial distribution (obtained using similar trick):

$$\sigma^2 = \frac{\sum_{m=0}^N (m - \mu)^2 P(m, N, p)}{\sum_{m=0}^N P(m, N, p)} = Npq$$

- ★ Example: Suppose you observed  $m$  special events (success) in a sample of  $N$  events
  - ◆ measured probability (“efficiency”) for a special event to occur:

$$\varepsilon = \frac{m}{N}$$

- ◆ error on the probability ("error on the efficiency"):

$$\sigma_\varepsilon = \frac{\sigma_m}{N} = \frac{\sqrt{Npq}}{N} = \frac{\sqrt{N\varepsilon(1-\varepsilon)}}{N} = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N}}$$

👉 sample ( $N$ ) should be as large as possible to reduce uncertainty in the probability measurement

- ★ Example: Suppose a baseball player's batting average is 0.33 (1 for 3 on average).
    - ◆ Consider the case where the player either gets a hit or makes an out (forget about walks here!).
      - probability for a hit:  $p = 0.33$
      - probability for “no hit”:  $q = 1 - p = 0.67$
    - ◆ On average how many hits does the player get in 100 at bats?
      - $\mu = Np = 100 \cdot 0.33 = 33$  hits
    - ◆ What's the standard deviation for the number of hits in 100 at bats?
      - $\sigma = (Npq)^{1/2} = (100 \cdot 0.33 \cdot 0.67)^{1/2} \approx 4.7$  hits
- 👉 we expect  $\approx 33 \pm 5$  hits per 100 at bats

## Poisson Probability Distribution

- A widely used discrete probability distribution
- Consider the following conditions:
  - ★  $p$  is very small and approaches 0
    - ◆ example: a 100 sided dice instead of a 6 sided dice,  $p = 1/100$  instead of  $1/6$
    - ◆ example: a 1000 sided dice,  $p = 1/1000$
  - ★  $N$  is very large and approaches  $\infty$ 
    - ◆ example: throwing 100 or 1000 dice instead of 2 dice
  - ★ product  $Np$  is finite
- Example: radioactive decay
  - ★ Suppose we have 25 mg of an element
    - ☞ very large number of atoms:  $N \approx 10^{20}$
  - ★ Suppose the lifetime of this element  $\lambda = 10^{12}$  years  $\approx 5 \times 10^{19}$  seconds
    - ☞ probability of a given nucleus to decay in one second is very small:  $p = 1/\lambda = 2 \times 10^{-20}/\text{sec}$
    - ☞  $Np = 2/\text{sec}$  **finite!**
    - ☞ number of counts in a time interval is a Poisson process
- Poisson distribution can be derived by taking the appropriate limits of the binomial distribution

$$P(m, N, p) = \frac{N!}{m!(N-m)!} p^m q^{N-m}$$

$$\frac{N!}{(N-m)!} = \frac{N(N-1)\cdots(N-m+1)(N-m)!}{(N-m)!} = N^m$$

$$q^{N-m} = (1-p)^{N-m} = 1 - p(N-m) + \frac{p^2(N-m)(N-m-1)}{2!} + \cdots \approx 1 - pN + \frac{(pN)^2}{2!} + \cdots \approx e^{-pN}$$



Siméon Denis Poisson  
June 21, 1781-April 25, 1840

$$P(m, N, p) = \frac{N^m}{m!} p^m e^{-pN}$$

Let  $\mu = Np$

$$P(m, \mu) = \frac{e^{-\mu} \mu^m}{m!}$$

$$\sum_{m=0}^{\infty} \frac{e^{-\mu} \mu^m}{m!} = e^{-\mu} \sum_{m=0}^{\infty} \frac{\mu^m}{m!} = e^{-\mu} e^{\mu} = 1$$

Poisson distribution is normalized

- ◆  $m$  is always an integer  $\geq 0$
- ◆  $\mu$  does **not** have to be an integer

★ It is easy to show that:

$\mu = Np = \text{mean of a Poisson distribution}$

$\sigma^2 = Np = \mu = \text{variance of a Poisson distribution}$

mean and variance are the same number

● Radioactivity example with an average of 2 decays/sec:

★ What's the probability of zero decays in one second?

$$p(0, 2) = \frac{e^{-2} 2^0}{0!} = \frac{e^{-2} \cdot 1}{1} = e^{-2} = 0.135 \rightarrow 13.5\%$$

★ What's the probability of more than one decay in one second?

$$p(> 1, 2) = 1 - p(0, 2) - p(1, 2) = 1 - \frac{e^{-2} 2^0}{0!} - \frac{e^{-2} 2^1}{1!} = 1 - e^{-2} - 2e^{-2} = 0.594 \rightarrow 59.4\%$$

★ Estimate the most probable number of decays/sec?

$$\left. \frac{\partial}{\partial m} P(m, \mu) \right|_{m^*} = 0$$



- ◆ To solve this problem its convenient to maximize  $\ln P(m, \mu)$  instead of  $P(m, \mu)$ .

$$\ln P(m, \mu) = \ln \left( \frac{e^{-\mu} \mu^m}{m!} \right) = -\mu + m \ln \mu - \ln m!$$

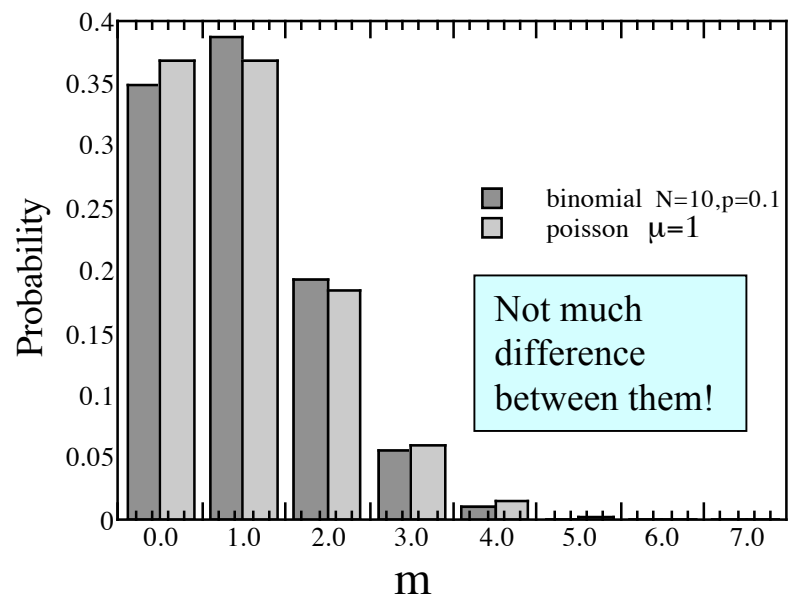
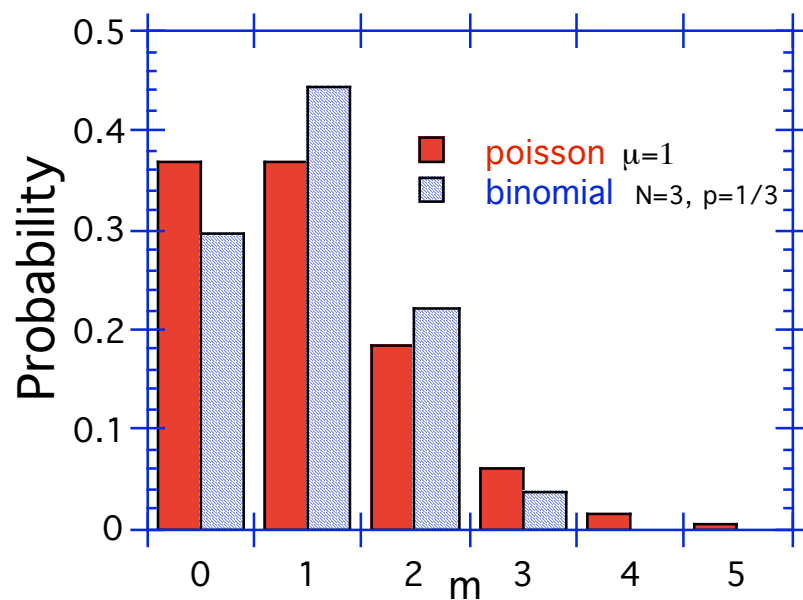
- ◆ In order to handle the factorial when take the derivative we use *Stirling's Approximation*:  
 $\ln m! \approx m \ln m - m$

$$\begin{aligned} \frac{\partial}{\partial m} \ln P(m, \mu) &= \frac{\partial}{\partial m} (-\mu + m \ln \mu - \ln m!) \\ &\approx \frac{\partial}{\partial m} (-\mu + m \ln \mu - m \ln m + m) \\ &= \ln \mu - \ln m - m \frac{1}{m} + 1 \\ &= 0 \end{aligned}$$

$$m^* = \mu$$

- 👉 The most probable value for  $m$  is just the average of the distribution
- 👉 If you observed  $m$  events in an experiment, the error on  $m$  is  $\sigma = \sqrt{\mu} = \sqrt{m}$
- ◆ This is only approximate since Stirlings Approximation is only valid for large  $m$ .
- ◆ Strictly speaking  $m$  can only take on integer values while  $\mu$  is not restricted to be an integer.

## Comparison of Binomial and Poisson distributions with mean $\mu = 1$



For large N: Binomial distribution looks like a Poisson of the same mean