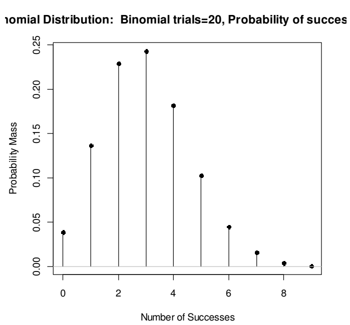
# Discrete Probability Distribution

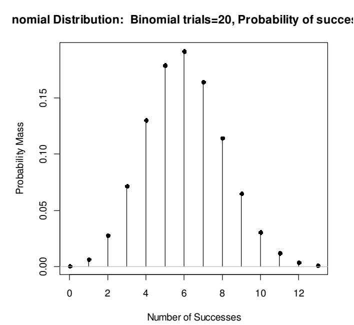
Week 11

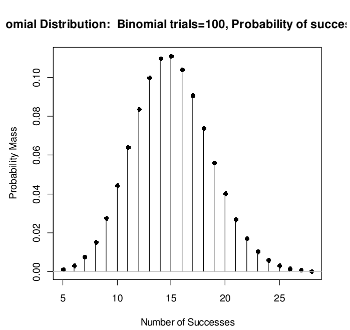
## Shapes of binomial distributions

For this activity you will use a web applet. Go to http://socr.stat.ucla.edu/htmls/SOCR_Experiments.html and choose Binomial coin experiment in the drop down menu on the left.

- Set the number of trials to 20 and the probability of success to 0.15. Describe the shape of the distribution of number of successes.
- What happens to the shape of the distribution as *n* stays constant and *p* changes?
- What happens to the shape of the distribution as *p* stays constant and *n* changes?

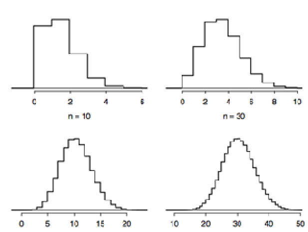



## Distributions of number of successes

Histograms of samples from the binomial model where $p = 0.10$ and $n = 10, 30, 100,$ and $300$. What happens as *n* increases?



## How large is large enough?

The sample size is considered large enough if the expected number of successes and failures are both at least 10.

$np \geq 10$ and $n(1 - p) \geq 10$

## How large is large enough?

The sample size is considered large enough if the expected number of successes and failures are both at least 10.

$np \geq 10$ and $n(1 - p) \geq 10$

10 x 0.13 ≈ 1.3

10 x (1 - 0.13) = 8.7

## Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

*A.* $n = 100$, $p = 0.95$
*B.* $n = 25$, $p = 0.45$
*C.* $n = 150$, $p = 0.05$
*D.* $n = 500$, $p = 0.015$

## Practice

Below are four pairs of Binomial distribution parameters. Which distribution can be approximated by the normal distribution?

*A.* $n = 100$, $p = 0.95$
*B.* *n = 25, p = 0.45 → 25 x 0.45 = 11.25, 25 x 0.55 = 13.75*
*C.* $n = 150$, $p = 0.05$
*D.* $n = 500$, $p = 0.015$

## Hypergeometric Distribution

- “n” trials in a sample taken from a finite population of size N.
- Samples taken without replacement
- Outcome & trials are dependent
- Concerned with finding the probability of “X” success in the sample where there are “S” successes in the population

## Hypergeometric Distribution

$$P(X) = \frac{C_X^S C_{n-X}^{N-S}}{C_n^N} = \frac{\frac{S!}{X!\,(S-X)!}\,\frac{(N-S)!}{(n-X)!\,(N-S-n+X)!}}{\frac{N!}{n!\,(N-n)!}}$$

where
N = population size
S = # of success in the population
N-S = # of failure in the population
n = sample size
X = # of successes in the sample
n-X = # of failures in the sample.

- $C_X^S$ : # of possible ways that X success can be selected for the sample out of S successes contained in the population.
- $C_{n-X}^{N-S}$ : (N-S) failures can be selected for the sample out of (N-S) failures in the population.
- $C_n^N$: total number of different sample size n that can be obtained from the population size N.

**Practice** $P(X) = \frac{C_X^S C_{n-X}^{N-S}}{C_n^N}$

3 different computers are checked from 10 in the department. 4 of the 10 computers have illegal software loaded. What is the probability that 2 of the 3 selected computers have illegal software loaded?

N=10
S=4
n=3
X=2

$$P(X) = \frac{C_2^4 C_{3-2}^{10-4}}{C_3^{10}}$$

$$= \frac{\frac{4!}{2!\,(4-2)!}\frac{(10-4)!}{(3-2)!\,(10-4-3+2)!}}{\frac{10!}{3!\,(10-3)!}}$$

$$= \frac{\frac{4!}{2!\,2!}\frac{6!}{1!\,5!}}{\frac{10!}{3!\,7!}} = \frac{6*6}{120} = 0.3$$

Probability of selecting 2 of the 3 selected computers have illegal software loaded is 30%.

**Practice**

A company receives a shipment of 16 items. A random sample of 4 items is selected and the shipment is rejected if any of these items proves to be defective.

Q1. What is the probability of accepting a shipment containing 4 defective items?

N=16, S=4, n=4, X=0
P(X=0)= 0.2720

Q2. What is the probability of accepting a shipment containing 1 defective item?

N=16, S=1, n=4, x=0
P(X=0)=0.75

Q3. What is the probability of rejecting a shipment containing 1 defective item?

N=16, S=1, n=4, X= ?

1-P(X=0)
=1-0.75
=0.25

## Poisson Distribution

Apply the Poisson Distribution when

- You wish to count the number of times an event occurs in a given continuous interval.
- The probability that an event occurs in one subinterval is very small and is the same for all subintervals.
- The number of events that occur in one subinterval is independent of the number of events that occur in the other subintervals.

## Poisson Distribution

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

- X: # of successes over a give time/space
- $\lambda$ : expected number of successes per given time/space.

Should be the same unit!

## Poisson Distribution

- Mean
- Variance

$\mu = E(X) = \lambda$ $\quad\quad$ $\sigma^2 = E[(x - \mu)^2] = \lambda$

## Practice

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

3 component failures occurred during last 100 days in his computer system.

a. What is the probability of no failure in a given day?

What is $\lambda$ ? $\lambda = 3/100 = 0.03$ per day.

$$P(X = 0|\lambda = 0.03)$$
$$= \frac{e^{-0.03}0.03^0}{0!} = e^{-0.03}$$
$$= 0.9704$$

## Practice

b. What is the probability of one or more components failures in a given day?

1-P(X=0|$\lambda$=0.03) =0.029554

c. What is the probability of at least 2 failures in a 3-day period?

What is $\lambda$ ?

$\lambda = 0.03*3 = 0.09/3$ days

$$P(X \geq 2|\lambda = 0.09)$$
$$= 1 - P(X \leq 1|\lambda = 0.09)$$
$$= 1 - [P(X = 0) + P(X = 1)$$
$$= 1 - [\frac{e^{-0.09}0.09^0}{0!} + \frac{e^{-0.09}0.09^1}{1!}]$$
$$= 1 - (e^{-0.09}(1 + 0.09))$$
$$= 0.0038$$

## Practice

- Customers arrive at a photocopy center at an average rate of two every 5 minutes.
- Find a probability that more than two customers arrive in a 5-minute period.

X: # of arriving customers in a 5 minutes period.

What is $\lambda$ ?

$\lambda = 2$ per 5 minutes.

What kind of probability are we looking for?

$P(X>2|\lambda = 2)$.

$$P(X > 2|\lambda = 2)$$
$$= 1 - P(X \leq 2|\lambda = 2)$$
$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$
$$= 1 - [\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}]$$
$$= 1 - (e^{-2}(1 + 2 + 2))$$
$$= 0.3233$$

## Practice'

- Customers arrive at a photocopy center at an average rate of two every 5 minutes.
- Find a probability that more than two customers arrive in a 1-minute period.

X: # of arriving customers in a 1 minutes period.

What is λ ?
λ =2/5.
What kind of probability are we looking for?
P(X>2|λ = 2/5).

## Practice

- A professor receives on average 4.2 e-mails from students the day before a final exam. If the distribution of calls is Poisson, what is the probability of receiving at least 3 of these e-mails on such a day?

What is λ? = 4.2
P(X≥3|λ=4.2)=1-[P(X=2)+P(X=1)+P(X=0)]
=1-e^-4.2[1+4.2+(4.2^2)/2] = 0.78976

## Poisson Approximation to the Binomial Distribution

- When n is very large and p is small (np≤ 7), then we can use Poisson distribution to approximate binomial distribution.

- X: number of successes from n independent trials
- P: probability of success
- Distribution of X: binomial with mean np.
- n: large, np≤ 7.

- This binomial distribution can be approximated by the Poisson distribution with λ = np.

## Poisson Approximation to Binomial Distribution (np≤ 7)

$$P(X) = \frac{e^{-np}(np)^x}{x!}$$

## Practice

- 3.5% of small corporate would file for bankruptcy in the coming year. For a random sample of 100 small corporations, what is the probability that at least 3 will file for bankruptcy in the next year?

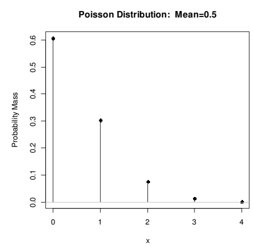
- n= 100, p = 0.035, np = 3.5
- P(X≥3| np=3.5)=1-P(X≤2|np=3.5)
- =1-[P(X=0)+P(X=1)+P(X=2)]
- =1- e^(-3.5)*[1+3.5+(3.5^2)/2]=0.684093

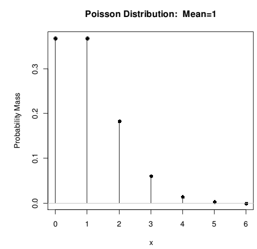
## If you try to solve this problem using Binomial Distribution…

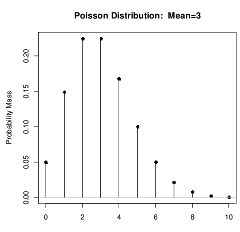
- $P(X \geq 3 | n=100, p=0.035) = 1 - P(X \leq 2 | n=100, p=0.035)$
- $= 1 - [P(X=0) + P(X=1) + P(X=2)]$
- $= 1 - \left[ \frac{100!}{0!\,100!}(0.035^0)(0.965^{100}) + \frac{100!}{1!\,99!}(0.035^1)(0.965^{99}) + \frac{100!}{2!\,98!}(0.035^2)(0.965^{98}) \right]$
- $= 0.6841$

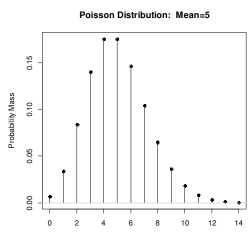
## Graphs of Poisson Distribution

- $\lambda = 0.5$ $\lambda = 1$




## Joint Probability Function

- A joint probability function is used to express the probability that X takes the specific value x and simultaneously Y takes the value y, as a function of x and y.
- $P(x,y) = P(X=x \cap Y=y)$

## Practice

- Coin toss – 3 times
- X: # of heads on the 1st toss
- Y: total number of heads
- S={hhh, hht, hth, htt, thh, tht, tth, ttt}
- Complete the table below, joint probabilities and marginal probabilities

| | | Y (total number of heads in 3 tosses) | | | | P(X) |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | |
| X (# of heads on 1st toss) | 0 | | | | | |
| | 1 | | | | | |
| P(Y) | | | | | | |

| | | Y (total number of heads in 3 tosses) | | | | P(X) |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | |
| X (# of heads on 1st toss) | 0 | {ttt} =1/8 | {tht, tth} =2/8 | {thh} =1/8 | 0 | =4/8 |
| | 1 | 0 | {htt} =1/8 | {hht, hth} =2/8 | {hhh} =1/8 | =4/8 |
| P(Y) | | =1/8 | =3/8 | =3/8 | =1/8 | |

## Conditional Probability Function

- The conditional probability function of the random variable Y expresses the probability that Y takes the value y when the value x is specified for X.

$$P(y|x) = \frac{P(x,y)}{P(x)} \qquad P(x|y) = \frac{P(x,y)}{P(y)}$$

| | | Y (total number of heads in 3 tosses) | | | | P(X) |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | |
| X (# of heads on 1st toss) | 0 | {ttt} =1/8 | {tht, tth} =2/8 | {thh} =1/8 | 0 | P(X=0)= 4/8 |
| | 1 | 0 | {htt} =1/8 | {hht, hth} =2/8 | {hhh} =1/8 | P(X=1)= 4/8 |
| P(Y) | | P(Y=0) =1/8 | P(Y=1) =3/8 | P(Y=2) =3/8 | P(Y=3)= 1/8 | |

- Find P(X=1|Y=2) and P(Y=2|X=1)

$$P(x|y) = \frac{P(x,y)}{P(y)} \qquad P(y|x) = \frac{P(x,y)}{P(x)}$$

## Quiz 8

a. 30% of cars on Eskisehir yolu are driving over the speed limit (82km/hr). What is the probability that at least 2 cars are going over the speed limit if there are 10 cars on the road? Calculate by hand (without table).

b. If 75% of cars on Eskisehir yolu are driving over the speed limit, what is the probability that more than 5 cars are going over the speed limit if there are 15 cars on the road? (Use table)

**TABLE II:** *continued*

| | | *p* | | | | | |
|---|---|---|---|---|---|---|---|
| *n* | *x* | 0.05 | 0.10 | 0.15 | 0.20 | 0.25 | 0.30 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 15 | 0 | 0.4633 | 0.2059 | 0.0874 | 0.0352 | 0.0134 | 0.0047 |
| | 1 | 0.8290 | 0.5490 | 0.3186 | 0.1671 | 0.0802 | 0.0353 |
| | 2 | 0.9638 | 0.8159 | 0.6042 | 0.3980 | 0.2361 | 0.1268 |
| | 3 | 0.9945 | 0.9444 | 0.8227 | 0.6482 | 0.4613 | 0.2969 |
| | 4 | 0.9994 | 0.9873 | 0.9383 | 0.8358 | 0.6865 | 0.5155 |
| | 5 | 0.9999 | 0.9978 | 0.9832 | 0.9389 | 0.8516 | 0.7216 |
| | 6 | 1.0000 | 0.9997 | 0.9964 | 0.9819 | 0.9434 | 0.8689 |
| | 7 | 1.0000 | 1.0000 | 0.9994 | 0.9958 | 0.9827 | 0.9500 |
| | 8 | 1.0000 | 1.0000 | 0.9999 | 0.9992 | 0.9958 | 0.9848 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9992 | 0.9963 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9993 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 |
| | 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| | 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

## Quiz 8_2

a. 40% of student taking ECO239 pass the course. What is the probability that at least 9 students will pass the course if there are 10 students in the class. Calculate by hand (without table).

b. If 70% of students taking ECO239 pass the course, what is the probability that more than 10 students will pass the course if there are 15 students in the class? Use table.