05_Binomial_Random_Variables

Bernoulli Random Variable

- Bernoulli trials are random events with three characteristics:
    - Two possible outcomes (success or failure, 0 or 1, etc)
    - Fixed probability of success for each outcome
    - Variables are independent
- Definition
    - A random variable *B* with two possible values
    - 1 = success and 0 = failure
    - E(*B*)=p
        - Var(*B*)=p(1-p) ## Random Variables for Counts
- The sum of independent and identical-distribution (IID) Bernoulli random variables is *Y*
- *Y*
    - Number of successes in n Bernoulli trials
    - Defined by parameters (n) and (p)
        - Each trial (n) has probability of success (p)
- **Properties of Binomial Random Variables**
    - Mean and Variance
        - E(Y) = np
            - Number of trials * the probability of each trial
        - Var(Y) = np(1-p)
            - Variance of random variables p(1-p) times number of variables
    - Consists of two parts:
        - The number of sequences that have *Y* successes in n attempts
        - The probability of a specific sequence of Bernoulli trials with *Y* success in n attempts
    - Binomial probability for *Y* success in n trials
        - $P(Y = y)= nC_y p^Y (1-p)^{n-y}$
        - nC_y: “n choose y”
            - Calculated (n!/(y!(n-y!))
        - Range of random variable *Y* is 0-n because you can have zero success to at most, n success /Images/05_BiProb_Formula.png

Binomial Probability for Y success in n trials

$$P(Y=y) = {}_nC_y \; p^y (1-p)^{n-y}$$

$${}_nC_y = \text{"n choose y"} = \frac{n!}{y!(n-y)!}$$

or

$$\binom{n}{y}$$

E.g. Probability of seeing 8 doctors in 10 visits.

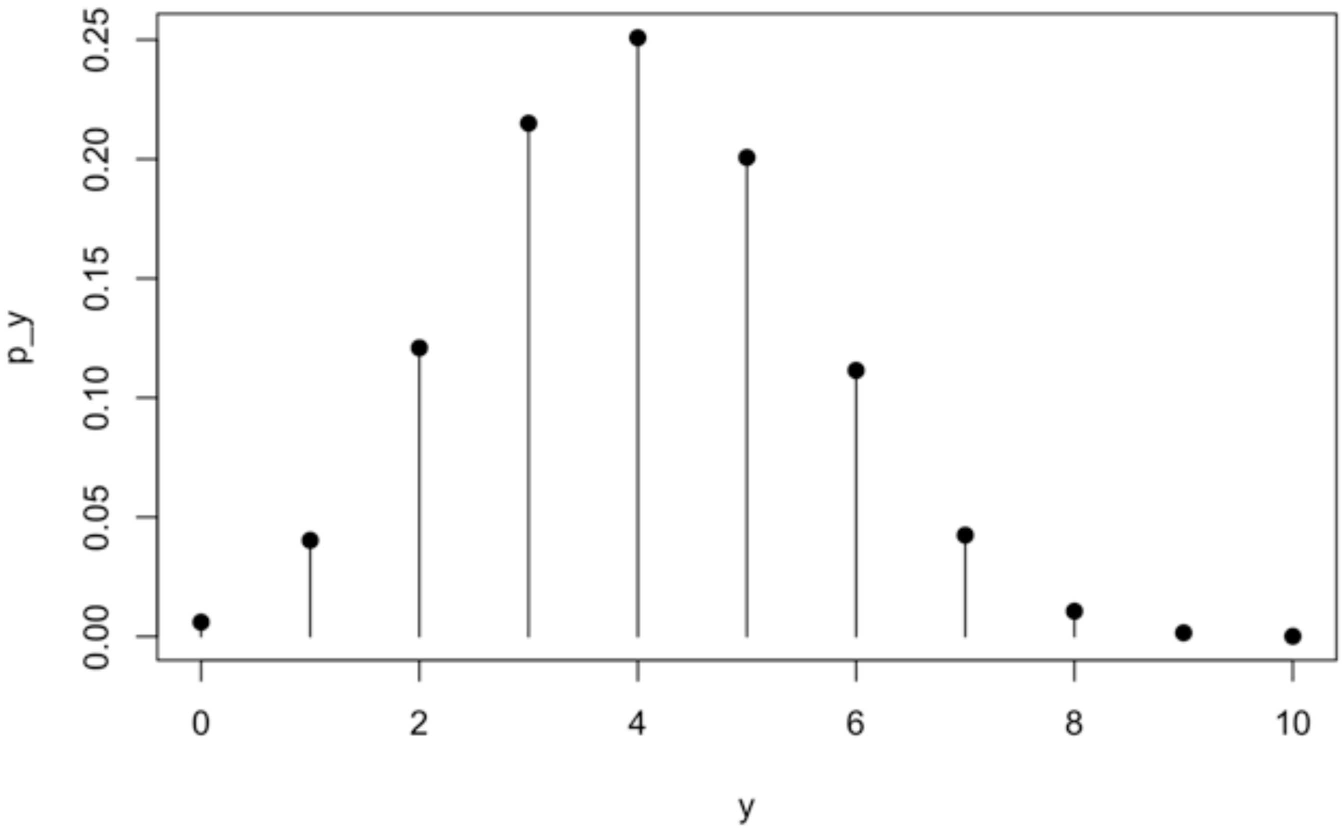
$$P(Y=8) = {}_{10}C_8 (0.4)^8 (0.6)^2 = 0.011 \approx 1\%$$

R Examples

```
y <- 0:10
# Discrete binomial with n of 10 and probability of .4
p_y <- dbinom(y, size = 10, prob = 0.4)

# Create distribution of y and probability of y
dist <- cbind(y,p_y)

# Name columns
colnames(dist) <- c("y", "p(y)")

# Call distribution
dist
```

```
##        y         p(y)
##  [1,]  0 0.0060466176
##  [2,]  1 0.0403107840
##  [3,]  2 0.1209323520
##  [4,]  3 0.2149908480
##  [5,]  4 0.2508226560
##  [6,]  5 0.2006581248
##  [7,]  6 0.1114767360
##  [8,]  7 0.0424673280
##  [9,]  8 0.0106168320
## [10,]  9 0.0015728640
## [11,] 10 0.0001048576
```

```
# Create plot with type = histogram
plot(y, p_y, type = "h")
points(y, p_y, pch=19)
```




```
# Cumulative probability
# p(Y>=8) = 1-P(Y<=7)
1-pbinom(7, size = 10, prob = 0.4)
```

```
## [1] 0.01229455
```

Example

/Images/05_BiProb_Example.png

Focus on Sales

A focus group with nine randomly chosen participants was shown a prototype of a new product and asked if they would buy it at a price of $99.95. Six of them said yes. The development team claimed that 80% of customers would buy the new product at that price. If the claim is correct, what results would we expect from the focus group?

Method

Use the binomial model for this situation. Each focus group member has two possible responses: yes, no. We can use $Y \sim \text{Bi}(n = 9, p = 0.8)$ to represent the number of yes responses out of nine.

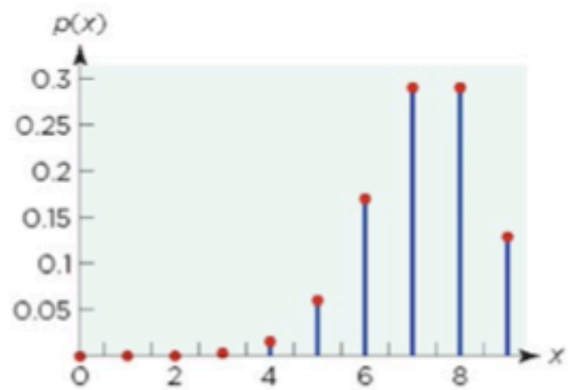
/Images/05_BiProb_Solution.png

Mechanics – Find $E(Y)$ and $SD(Y)$

$E(Y) = np = (9)(0.8) = 7.2$
$Var(Y) = np(1\text{-}p) = (9)(0.8)(0.2) = 1.44$
$SD(Y) = 1.2$

The expected number is higher than the observed number of 6.

Mechanics – Probability Distribution



P(Y=6) = 0.18. While 6 is not the most likely outcome, it is still common.

Results seem in line with development claims.

Poisson Distribution

- Measures counts in a continuous interval of trials
- Poisson Random Variable
  - Describes the number of events determined by a random process during an interval of time or space
  - Is not finite, possible values are infinite
    - Opposed to binomial where range is 0:n
- Example: How many people enter a store in a day?

- Represented by λ
    - The rate of events
- $P(X = x) = e^{\lambda}(\lambda^{x}/x!)$ /Images/05_Poisson_Formula.png
    - e is a constant, 2.71828
    - λ = rate or average number of counts in an interval of time
    - x = number of counts
        - P(x) = probability of number of counts in your range
- $E(X) = \lambda$
- $Var(X) = \lambda$ Example:

/Images/05_Poisson_Example.png

Motivation

A supplier claims that its wafers have 1 defect per 400 cm^2. Each wafer is 20 cm in diameter, so the area is 314 cm^2. What is the mean number of defects and the standard deviation?

The random variable is the number of defects on a randomly selected wafer. The Poisson model applies.

1 defect 400 cm^2

λ defects 314 cm^2

$$\lambda = \frac{314}{400} = 0.785$$

$$E(y) = \lambda = 0.785 \text{ defects}$$

$$Var(y) = \lambda = 0.785 \text{ defects}$$

$$SD(y) = \sqrt{0.785} = 0.886 \text{ defects}$$

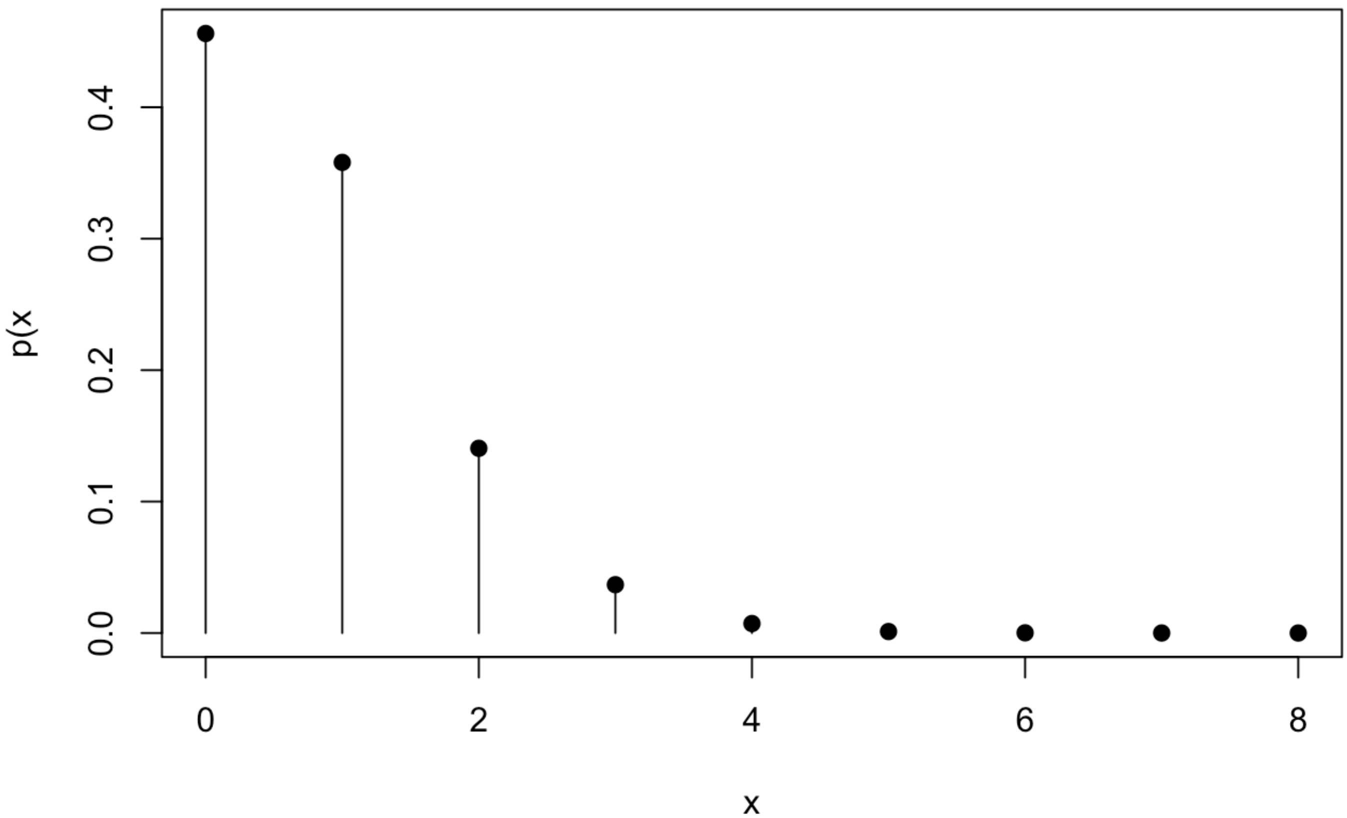
Probability of defect-free wafer

$$P(y=0) = \frac{e^{-0.785}\, 0.785^0}{0!} = 0.456 \approx 46\%$$

```
# X is number of defects per wafer
x <- 0:8

#Use dpois for probability of specific x
p_x <- dpois(x, lambda=314/400)

dist <- cbind(x,p_x)
colnames(dist) <- c("x", "p(x")
dist
```

```
##       x          p(x
##  [1,] 0 4.561197e-01
##  [2,] 1 3.580540e-01
##  [3,] 2 1.405362e-01
##  [4,] 3 3.677363e-02
##  [5,] 4 7.216826e-03
##  [6,] 5 1.133042e-03
##  [7,] 6 1.482396e-04
##  [8,] 7 1.662401e-05
##  [9,] 8 1.631231e-06
```

```
plot(dist,type="h")
points(dist, pch=19)
```




```
# P(x = 0)
dpois(0, lambda=314/400)
```

```
## [1] 0.4561197
```

```
# P(x>=3) = 1-P(x<=2)
# For cumulative use ppois instead of dpois
1-ppois(2, lambda=314/400)
```

```
## [1] 0.04529015
```

```
# Simulate random Poisson numbers with same lambda
set.seed(123)

# Use rpois for random
randPoisson <- rpois(100, lambda=314/400)
head(randPoisson)
```

```
## [1] 0 1 0 2 2 0
```

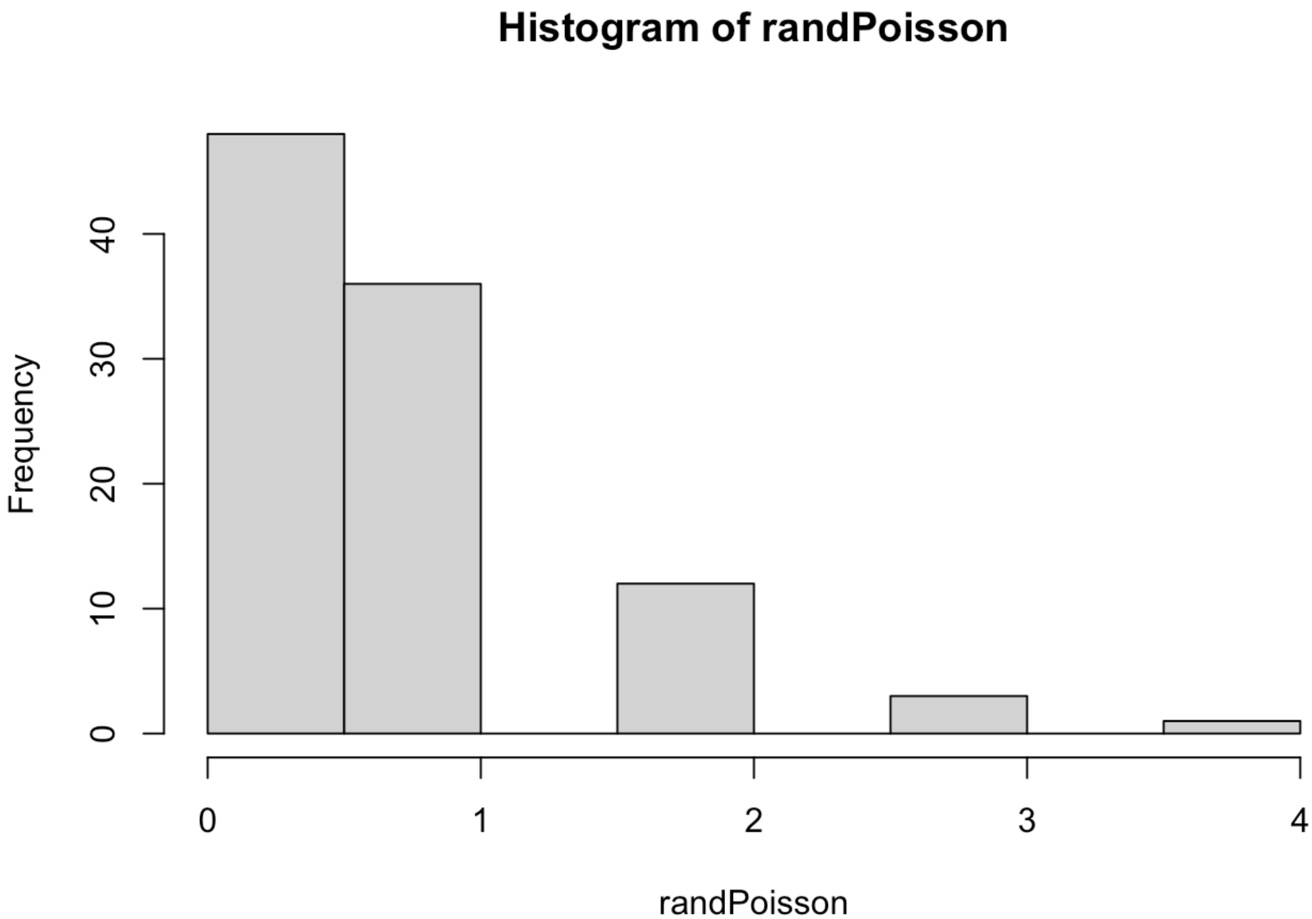
```
mean(randPoisson)
```

```
## [1] 0.73
```

```
var(randPoisson)
```

```
## [1] 0.7445455
```

```
hist(randPoisson)
```



Best Practices and Pitfalls

- Best practices
  - Ensure you have Bernoulli trials if you're going to use the binomial model
    - Check three rules
  - If trials are continuous, use Poisson random variable
- Pitfalls
  - Do not presume independence without checking
  - Do not assume stable conditions routinely