Discrete Probability Distributions

Random Variables

- Random Variable (RV): A numeric outcome that results from an experiment
- For each element of an experiment's sample space, the random variable can take on exactly one value
- Discrete Random Variable: An RV that can take on only a finite or countably infinite set of outcomes
- Continuous Random Variable: An RV that can take on any value along a continuum (but may be reported "discretely"
- Random Variables are denoted by upper case letters (Y)
- Individual outcomes for RV are denoted by lower case letters (y)

Probability Distributions

- Probability Distribution: Table, Graph, or Formula that describes values a random variable can take on, and its corresponding probability (discrete RV) or density (continuous RV)
- Discrete Probability Distribution: Assigns probabilities (masses) to the individual outcomes
- Continuous Probability Distribution: Assigns density at individual points, probability of ranges can be obtained by integrating density function
- Discrete Probabilities denoted by: $p(y) = P(Y=y)$
- Continuous Densities denoted by: $f(y)$
- Cumulative Distribution Function: $F(y) = P(Y \leq y)$

Discrete Probability Distributions

Probability (Mass) Function :

$$p(y) = P(Y = y)$$

$$p(y) \geq 0 \quad \forall y$$

$$\sum_{\text{all } y} p(y) = 1$$

Cumulative Distribution Function (CDF) :

$$F(y) = P(Y \leq y)$$

$$F(b) = P(Y \leq b) = \sum_{y=-\infty}^{b} p(y)$$

$$F(-\infty) = 0 \quad F(\infty) = 1$$

$F(y)$ is monotonically increasing in y

Example – Rolling 2 Dice (Red/Green)

Y = Sum of the up faces of the two die. Table gives value of *y* for all elements in *S*

Red\Green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

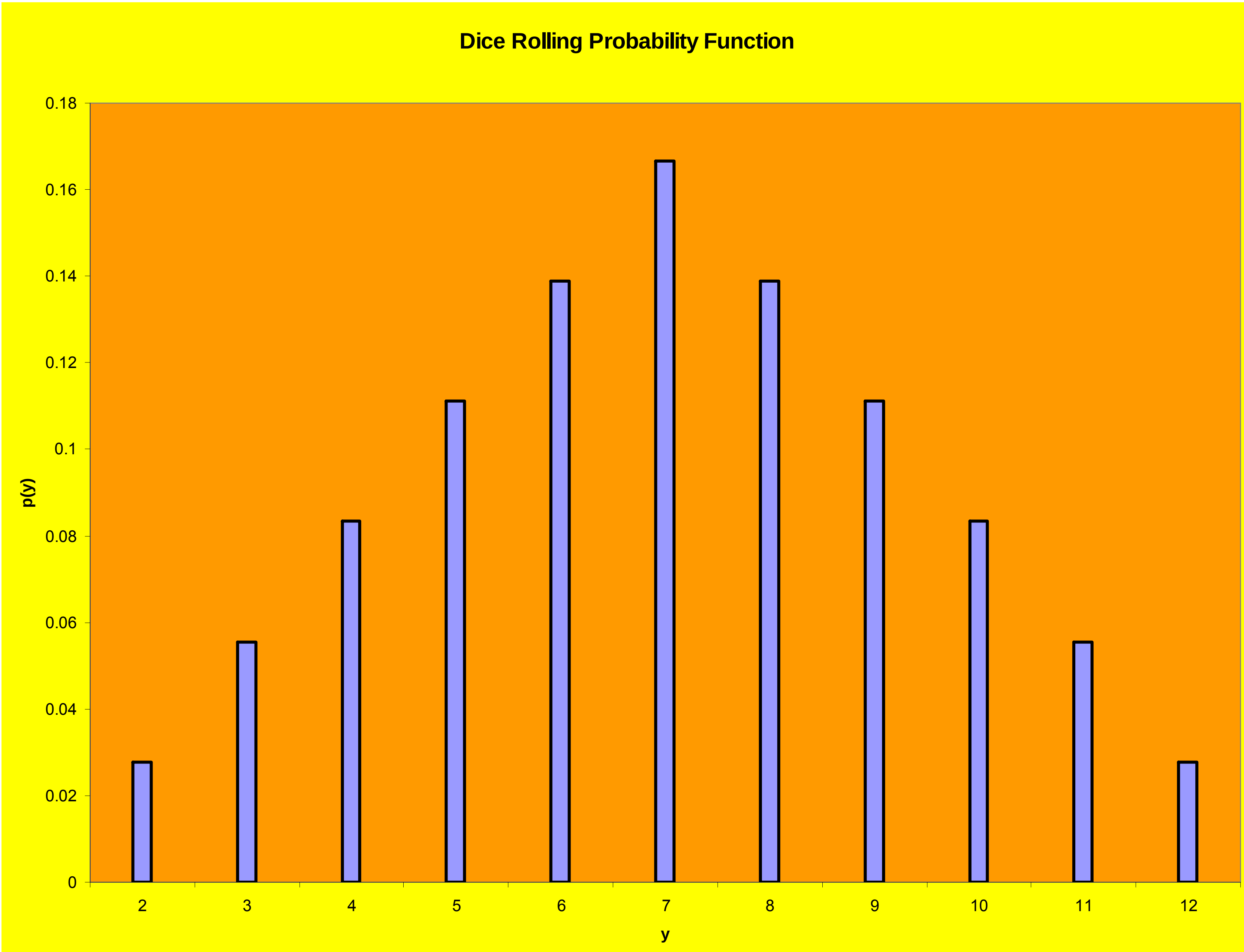
Rolling 2 Dice – Probability Mass Function & CDF

y	p(y)	F(y)
2	1/36	1/36
3	2/36	3/36
4	3/36	6/36
5	4/36	10/36
6	5/36	15/36
7	6/36	21/36
8	5/36	26/36
9	4/36	30/36
10	3/36	33/36
11	2/36	35/36
12	1/36	36/36

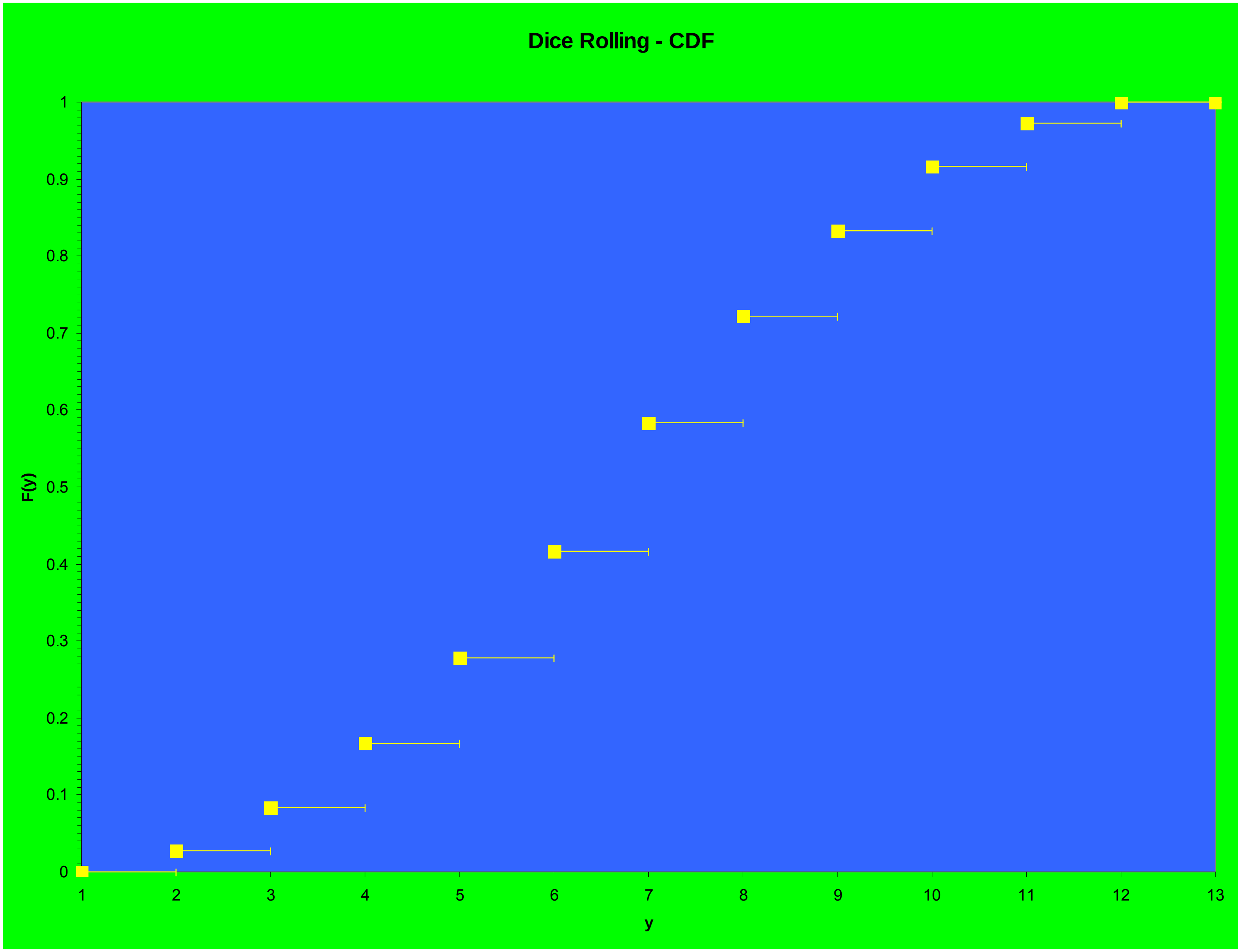
$$p(y) = \frac{\text{\# of ways 2 die can sum to } y}{\text{\# of ways 2 die can result in}}$$

$$F(y) = \sum_{t=2}^{y} p(t)$$

Rolling 2 Dice – Probability Mass Function



Rolling 2 Dice – Cumulative Distribution Function



Expected Values of Discrete RV's

- Mean (aka Expected Value) – Long-Run average value an RV (or function of RV) will take on
- Variance – Average squared deviation between a realization of an RV (or function of RV) and its mean
- Standard Deviation – Positive Square Root of Variance (in same units as the data)
- Notation:
  - Mean: $E(Y) = \mu$
  - Variance: $V(Y) = \sigma^2$
  - Standard Deviation: σ

Expected Values of Discrete RV's

$$\text{Mean}: \; E(Y) = \mu = \sum_{\text{all } y} y p(y)$$

$$\text{Mean of a function } g(Y): \quad E\left[g(Y)\right] = \sum_{\text{all } y} g(y) p(y)$$

$$\text{Variance}: \; V(Y) = \sigma^2 = E\left[(Y - E(Y))^2\right] = E\left[(Y - \mu)^2\right] =$$

$$= \sum_{\text{all } y} (y - \mu)^2 p(y) = \sum_{\text{all } y} \left(y^2 - 2y\mu + \mu^2\right) p(y) =$$

$$= \sum_{\text{all } y} y^2 p(y) - 2\mu \sum_{\text{all } y} y p(y) + \mu^2 \sum_{\text{all } y} p(y) =$$

$$= E\left[Y^2\right] - 2\mu(\mu) + \mu^2(1) = E\left[Y^2\right] - \mu^2$$

$$\text{Standard Deviation}: \sigma = +\sqrt{\sigma^2}$$

Expected Values of Linear Functions of Discrete RV's

$$\text{Linear Functions: } g(Y) = aY + b \quad (a, b \equiv \text{constants})$$

$$E[aY+b] = \sum_{\text{all } y} (ay+b)\,p(y) =$$

$$= a\sum_{\text{all } y} yp(y) + b\sum_{\text{all } y} p(y) = a\mu + b$$

$$V[aY+b] = \sum_{\text{all } y} \left((ay+b)-(a\mu+b)\right)^2 p(y) =$$

$$\sum_{\text{all } y} \left(ay - a\mu\right)^2 p(y) = \sum_{\text{all } y} \left[a^2\left(y-\mu\right)^2\right] p(y) =$$

$$= a^2 \sum_{\text{all } y} (y-\mu)^2\, p(y) = a^2\sigma^2$$

$$\sigma_{aY+b} = |a|\sigma$$

Example – Rolling 2 Dice

y	p(y)	yp(y)	y^2p(y)
2	1/36	2/36	4/36
3	2/36	6/36	18/36
4	3/36	12/36	48/36
5	4/36	20/36	100/36
6	5/36	30/36	180/36
7	6/36	42/36	294/36
8	5/36	40/36	320/36
9	4/36	36/36	324/36
10	3/36	30/36	300/36
11	2/36	22/36	242/36
12	1/36	12/36	144/36
Sum	36/36 =1.00	252/36 =7.00	1974/36= 54.833

$$\mu = E(Y) = \sum_{y=2}^{12} yp(y) = 7.0$$

$$\sigma^2 = E\left[Y^2\right] - \mu^2 = \sum_{y=2}^{12} y^2 p(y) - \mu^2$$

$$= 54.8333 - (7.0)^2 = 5.8333$$

$$\sigma = \sqrt{5.8333} = 2.4152$$

Binomial Experiment

- Experiment consists of a series of n identical trials
- Each trial can end in one of 2 outcomes: Success (S) or Failure (F)
- Trials are independent (outcome of one has no bearing on outcomes of others)
- Probability of Success, p, is constant for all trials
- Random Variable Y, is the number of Successes in the n trials is said to follow Binomial Distribution with parameters n and p
- Y can take on the values $y=0,1,\ldots,n$
- Notation: $Y \sim Bin(n,p)$

Binomial Distribution

Consider outcomes of an experiment with 3 Trials:

$SSS \Rightarrow y = 3 \quad P(SSS) = P(Y = 3) = p(3) = p^3$

$SSF, SFS, FSS \Rightarrow y = 2 \quad P(SSF \cup SFS \cup FSS) = P(Y = 2) = p(2) = 3p^2(1-p)$

$SFF, FSF, FFS \Rightarrow y = 1 \quad P(SFF \cup FSF \cup FFS) = P(Y = 1) = p(1) = 3p(1-p)^2$

$FFF \Rightarrow y = 0 \quad P(FFF) = P(Y = 0) = p(0) = (1-p)^3$

In General:

1) # of ways of arranging y S^s (and $(n-y)$ F^s) in a sequence of n positions $\equiv \binom{n}{y} = \frac{n!}{y!(n-y)!}$

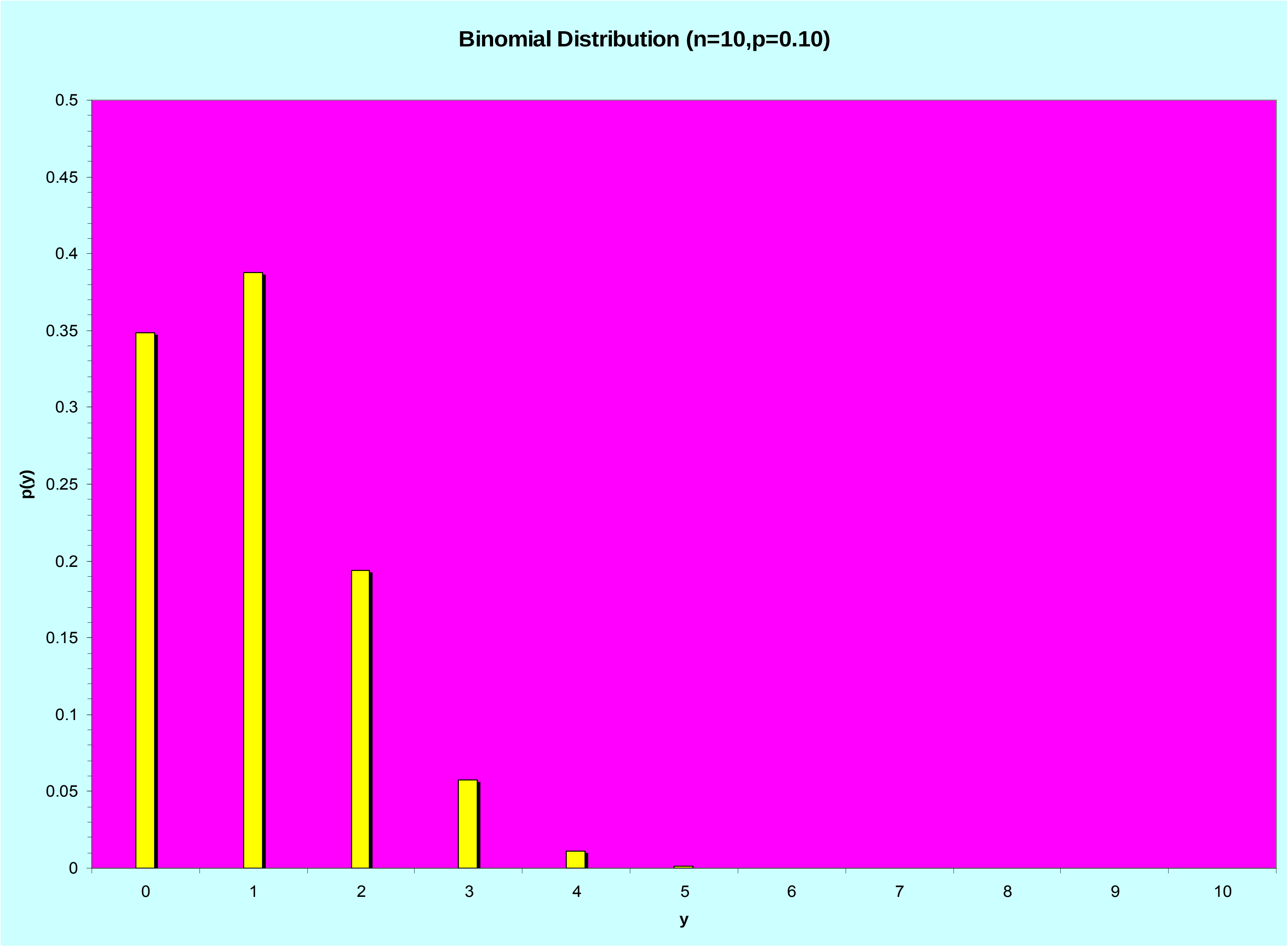
2) Probability of each arrangement of y S^s (and $(n-y)$ F^s) $\equiv p^y(1-p)^{n-y}$

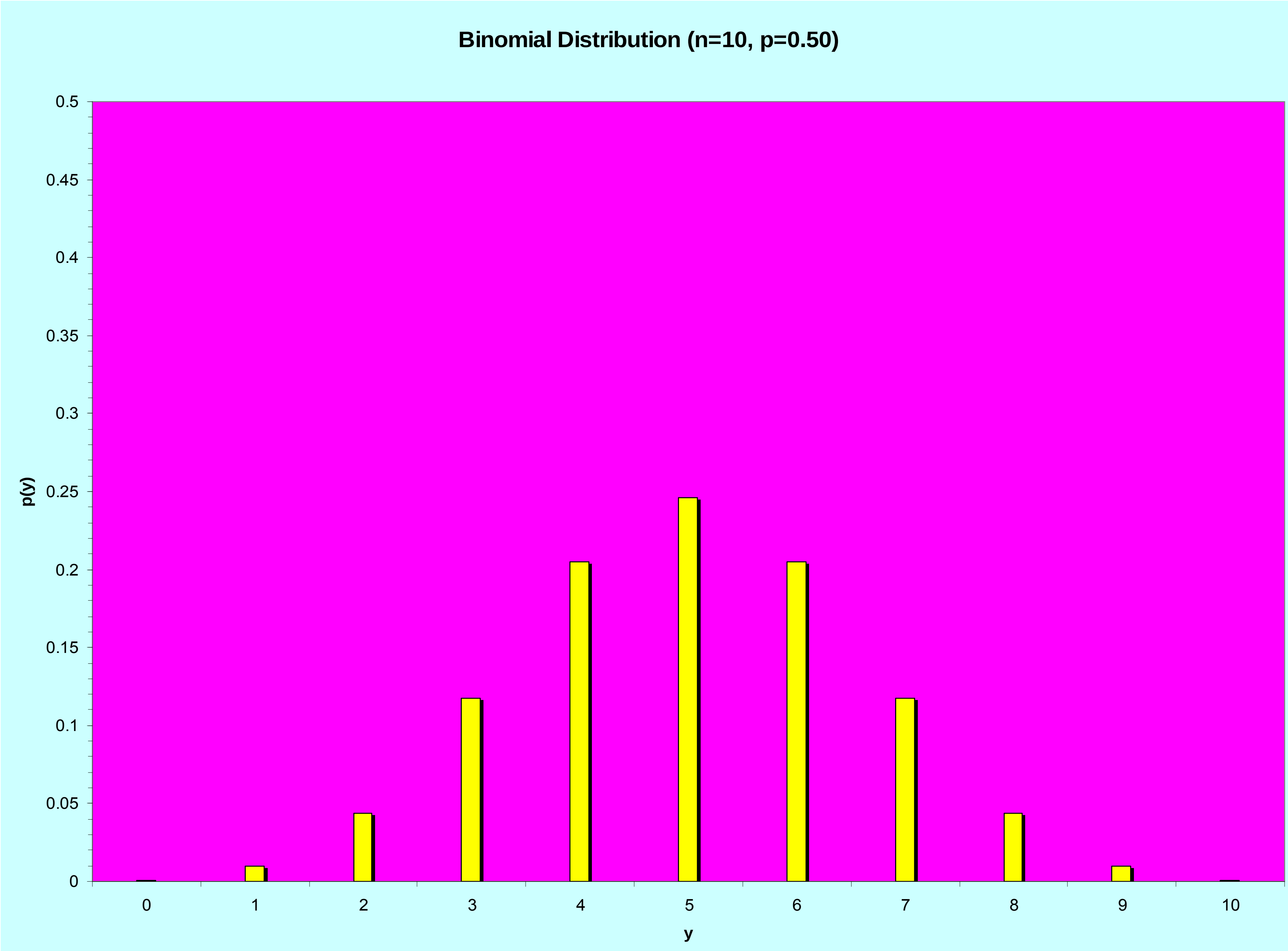
3) $\Rightarrow P(Y = y) = p(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad y = 0,1,\ldots,n$

EXCEL Functions:

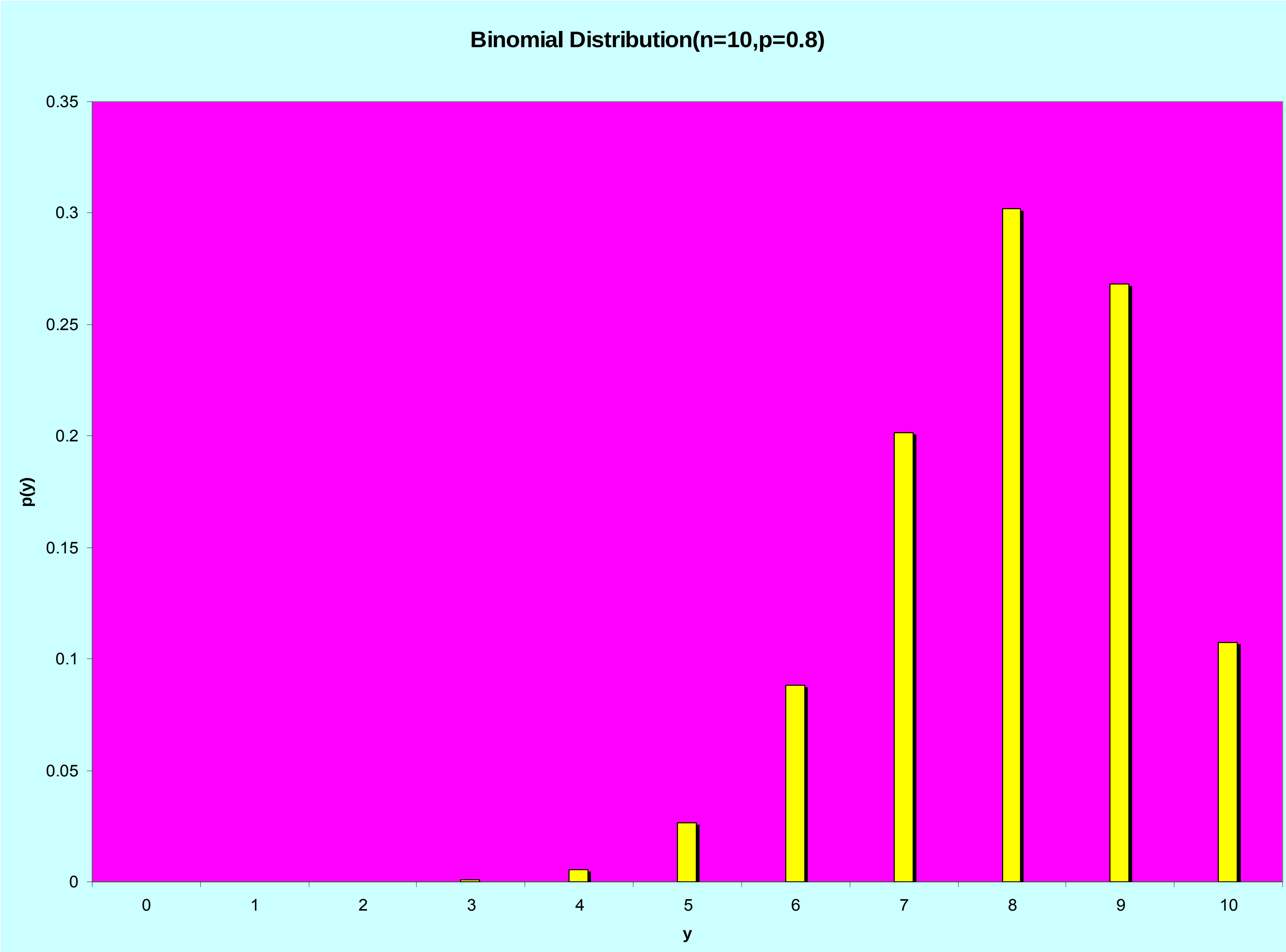
$p(y)$ is obtained by function: =BINOMDIST(y, n, p,0)

$F(y) = P(Y \leq y)$ is obtained by function: =BINOMDIST(y, n, p,1)


Binomial Distribution (n=10,p=0.10)
p(y)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10
y



Binomial Distribution (n=10, p=0.50)
p(y)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10
y



Binomial Distribution(n=10,p=0.8)
p(y)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10
y

Poisson Distribution

- Distribution often used to model the number of incidences of some characteristic in time or space:
  - Arrivals of customers in a queue
  - Numbers of flaws in a roll of fabric
  - Number of typos per page of text.
- Distribution obtained as follows:
  - Break down the "area" into many small "pieces" (n pieces)
  - Each "piece" can have only 0 or 1 occurrences (p=P(1))
  - Let $\lambda = np \equiv$ Average number of occurrences over "area"
  - $Y \equiv$ # occurrences in "area" is sum of 0^s & 1^s over "pieces"
  - $Y \sim \text{Bin}(n,p)$ with $p = \lambda/n$
  - Take limit of Binomial Distribution as $n \to \infty$ with $p = \lambda/n$

$$p(y) = P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad \lambda > 0, \quad y = 0,1,2,\ldots$$

Assignment

Q.1 Is the function defined as follows a density function?

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 & otherwise \end{cases}$$

if so, find $P(1 \leq X \leq 2)$

Q.2 In a Lottery, m tickets are drawn at a time out of n tickets numbered from 1 to n. find the expected value of the sum of the numbers on the tickets drawn.

Q.3 A cubical die is thrown 9000 times and a throw of 3 or 4 is observed 3240 times. Show that the die can't be regarded as an unbiased one and find the extreme limits between which the probability of a throw of 3 or 4 lies.

Q.4 The means of a simple samples of sizes 1000 and 2000 are 67.5 and 68.0 cm respectively. Can the sample be regarded as drawn from the same population of S.D. 2.5 cm?

Q.5 Define Student's-t-Distribution

Thank you