# Physics 307L

Spring 2021

Prof. Darcy Barron

Probability Distribution Functions

# Reminders

- Schedule of assignment due dates through the end of the semester is now posted on wiki, and in Teams
    - https://ghz.unm.edu/juniorlab/index.php?title=Schedule_Spring_2021#Course_Schedule
- **Schedule for Talk 2 is posted, first talks April 19, 26**
    - To get full credit, you need to submit your slides at least 1 hour ahead of time (submit through assignment in Teams)
    - See talk 2 scoring rubric (slide 3)
- **Lab report 2 due Wed, April 28 by noon**
    - Working on giving detailed feedback on lab report 1 this week
- **Lab report 3 is due Wed, May 12**
    - You must use LaTeX for this final lab report.
    - Ask if you need help finding and using an appropriate template
    - See resources in Lecture 5

# Talk 2 Rubric

- Expect everyone talk for at least 7 minutes and have at least 7 content slides
- Must send your slides at least one hour in advance of lecture start to get full credit
- Need to ask at least one question of another presenter to get full credit
- Score breakdown
  - Written Slides – 20%
  - Oral Presentation – 20%
  - Technical Content – 50%
  - Questions and Discussion – 10%

# AMERICAN ASTRONOMICAL SOCIETY

Enhancing and sharing humanity's scientific understanding of the universe since 1899.

## AAS Chambliss Student Achievement Awards Judging Form: Undergraduate Students

**Poster Number:** ____________________ **Student's Name:** ____________________________________________

**Judge's Name:** ________________________________________ **Judging Day (circle): Mon Tues Wed**

**Judging Times:** Morning breaks (9:30 am to 10:00 am) and again during the afternoon poster session (5:30 pm to 6:30 pm Mon., Tue., & Wed.; 1:00 pm to 2:00 pm Thu).

**If student is not present at his/her poster at either of the judging time, his/her poster is disqualified. Simply check this box ☐ and return this form to the AAS registration desk.**

**Directions:** Content is weighted by 2/3rd and Presentation by 1/3rd. Circle one underlined response in each of the brackets for EACH BULLET that best describes this student's poster and presentation. Enter the corresponding number (1, 2, 3, or 4) in the Score column. Sum the Score column and enter the score in the Total Score cell.

| Judging Criteria | | 4=Exemplary (accomplished); 3=Proficient (effective); 2=Basic (limited); 1=Below Basic (lacking) | Score |
|---|---|---|---|
| Student's Content (weighted by 2/3rd) | Conceptual Understanding of Research Within the Broader Context of Astronomy | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s) the outstanding question(s) or gap(s) in our understanding of astronomy related to their work. | ____ x 2 = ____ |
| | | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s) how their work *could potentially* help to answer these questions. | ____ x 2 = ____ |
| | Knowledge of How the Research was Conducted | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s) their methods of data collection and analysis, as well as their key physical assumptions. | ____ x 2 = ____ |
| | | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s), when appropriate, errors in and/or limitations of their methods of data collection and analysis, as well as their key physical assumptions. | ____ x 2 = ____ |
| | Understanding of the Results and Implications of the Research | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s), when appropriate, how their results and implications of their work *did* improve our understanding of astronomy. | ____ x 2 = ____ |
| | | • Student [ easily and concisely (4) / sufficiently (3) / is somewhat able to (2) / struggles to or cannot (1) ] describe(s), when appropriate, errors in and/or limitations of their results and implications of how their work *did not* help to improve our understanding of astronomy. | ____ x 2 = ____ |
| Student's Presentation (weighted by 1/3rd) | Poster Mechanics | • Poster [ easily and concisely (4) / sufficiently (3) / somewhat (2) / limits or prohibits (1) ] leads/leading Reader through a logical flow from (e.g.) title, to introduction, explanation of work, summary/conclusion, and references. | = ____ |
| | | • Graphics include [ all (4) / sufficiently (3) / some (2) / almost no (1) ] appropriate labels and units. | = ____ |
| | Verbal Organization of Research | • Student's oral presentation was [ extremely (4) / sufficiently (3) / somewhat (2) / limited in being or was not (1) ] clear, concise, and logical. | = ____ |
| | | • Listener could [ easily (4) / sufficiently (3) / somewhat (2) / was limited in or could not (1) ] follow(ing) lines of reasoning. | = ____ |
| | Verbal Interaction with Others | • Student was [ extremely (4) / sufficiently (3) / somewhat (2) / struggled to be or was/did not (1) ] articulate, use proper volume, use appropriate language for the Student's level of education/expertise, and convey a high level of confidence/poise. | = ____ |
| | | • Student was [ extremely (4) / sufficiently (3) / somewhat (2) / mostly not or not (1) ] consistent and adequate in responding to questions, including clarifying and restating as necessary. | = ____ |
| | | **TOTAL SCORE** | |

# Error Analysis

- Some steps in error analysis so far
  - Estimating uncertainties from equipment
  - Repeating measurements to estimate uncertainty
  - Propagating uncertainties
  - Plotting and correlating data
  - Choosing how to combine separate measurements, and possibly rejecting/cleaning data
  - Least-squares fitting with errors in both dimensions
- There are many techniques for fitting and analyzing your experimental data to understand its statistics and sources of **random/statistical error**

# What are statistics?

- A statistic summarizes data (data reduction)
- Statistics are the basis for using the data to make a decision
- Example: Is the faint smudge on an image a star or a galaxy?
  - Measure FWHM of the point-spread function.
  - Measure full-width-half-maximum, the FWHM.
  - The data set, the image of the object, is now represented by a *statistic*

**Median** value of a data ensemble $m_{1/2}$

Half of all data > $m_{1/2}$

Half of all data < $m_{1/2}$

---

**Deviation** of a data point about the **mean**: $d_i = x_i - \overline{x}$

**Average deviation:** $\overline{d} = \overline{x} - \overline{x} = 0$ *Not useful*

**Variance**: $\sigma^2 = \frac{1}{N-1} \sum_i^N d_i^2 = \frac{1}{N-1} \sum_i^N (x_i - \overline{x})^2$

**Standard deviation**: $\sqrt{\sigma}$

# What is statistical analysis?

- 1. Formulate a hypothesis
- 2. Gather data to test the hypothesis (via experiment, or by finding existing datasets)
- 3. Compare with the expected probability of that result (the sampling distribution)

Problems:

We don't know the actual underlying distribution

Small sample size

# Important uses of statistics

- Statistics can create precise statements for stating the logic of what we are doing and why
- Statistics allow us to quantify uncertainty
  - Measured quantities are basically useless without some measure of the associated range/error
  - Sometimes this can be inferred, but much better to be explicit (e.g. 5 photons, 72.1 degrees)
- Statistics help us avoid pitfalls like confirmation bias
- Statistics help make decisions about data

# PROBABILITY DISTRIBUTION FUNCTIONS

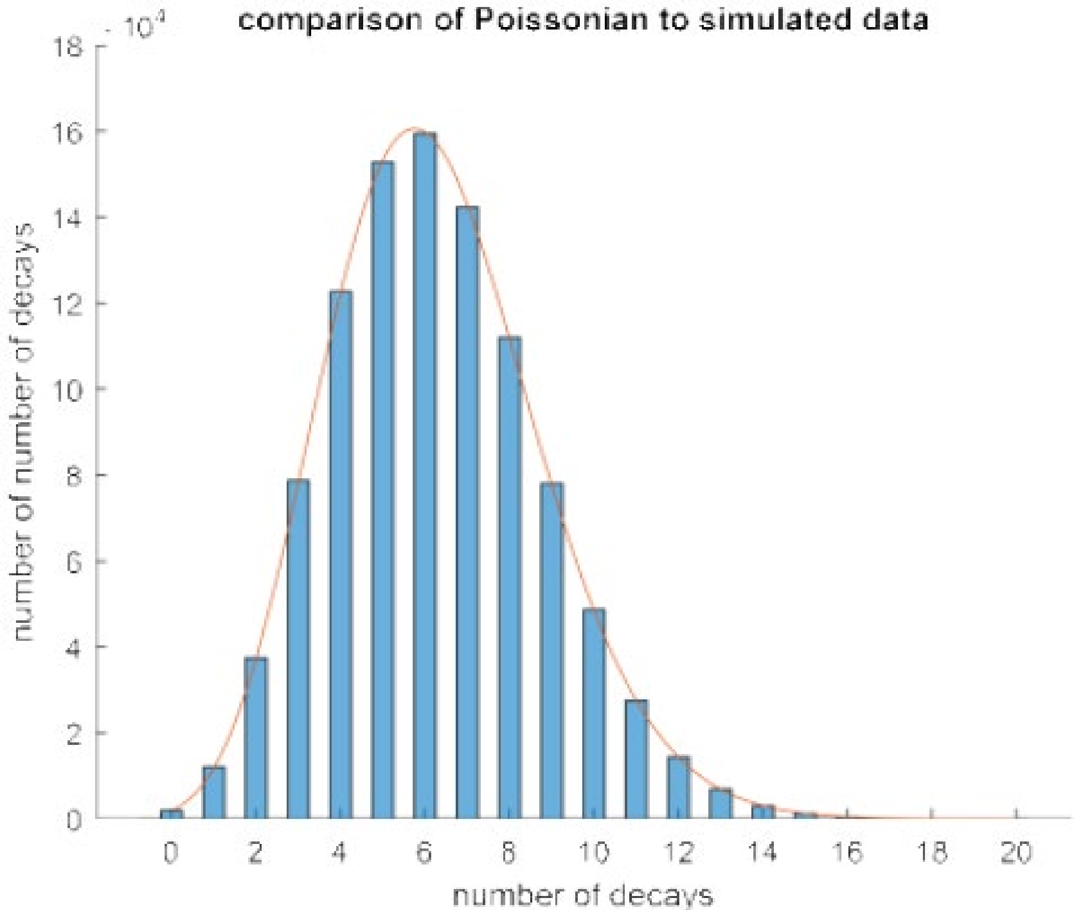
- **GAUSSIAN:** Random data, experimental parameters uncertain
  Described in Chapter 5 of Taylor

- **POISSON:** Number of counts in a specified time interval
  Described in Chapter 11 of Taylor

- **BINOMIAL:** Small number of possible outcomes (eg. heads or tails)
  Described in Chapter 10 of Taylor

These are models that may describe your data.

# Poisson Statistics Modeling



# PROBABILITY DISTRIBUTION FUNCTIONS

- **GAUSSIAN:** Random data, experimental parameters uncertain
- **POISSON:** Number of counts in a specified time interval
- **BINOMIAL:** Small number of possible outcomes (eg. heads or tails)

# BINOMIAL DISTRIBUTION: 1 COIN


OR


*p*=50%

*p*=50%

# BINOMIAL DISTRIBUTION: 2 COINS


***p*=25%**


***p*=25%**


***p*=25%**


***p*=25%**

# BINOMIAL DISTRIBUTION: 3 COINS


**8 DIFFERENT OUTCOMES: *p* = 12.5%**

# SAME COIN TOSSED 3 TIMES

**8 DIFFERENT OUTCOMES: *p* = 12.5%**

Possible states for the coin: $S = 2$

Number of coins flipped once: $N$

*– or –*

Number of times single coin is flipped: $N$

Possible outcomes = $S^N = 2^3 = 8$

Possible states for single die: *S = 6*

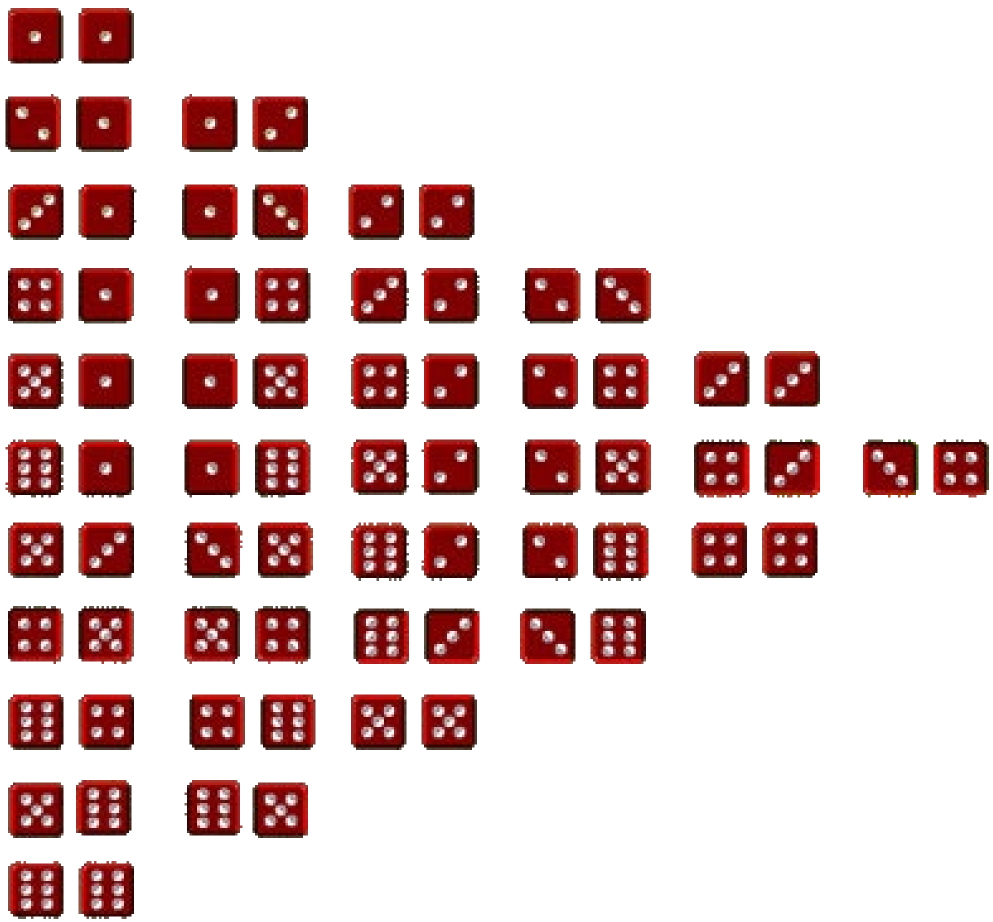
Number of times a single die is thrown: *N = 1*

Possible outcomes = $S^N = 6^1 = 6$

Possible states for single die: $S = 6$

Number of dice thrown: $N = 2$

Possible outcomes = $S^N = 6^2 = 36$



# PERMUTATIONS

Starters: $N = 6$


**6 different winners are possible**

# EXACTA: Pick the correct order of finish 1-2

Starters: $N = 6$


6 different winners are possible
Once winner is specified only five 2nd places possible
Number of different 1-2 sequences = 6 x 5 = **30**

# TRIFECTA: Pick the correct order of finish 1-2-3

Starters: $N = 6$


6 different winners are possible
Once winner is specified only five 2nd places possible; then four 3rd place finishes possible
Number of different 1-2-3 sequences = 6 x 5 x 4 = **120**

# Number of different race outcomes: 1-2-3-4-5-6


Starters: $N = 6$

6 x 5 x 4 x 3 x 2 x 1 = **720** different outcomes = $N!$

*P*: Number of possible **PERMUTATIONS**

*N*: Number of trials, events, participants, etc

*x*: Sequence of outcomes

$$P = \frac{N!}{(N - x)!}$$


Winner: $P = \frac{6!}{(6-1)!} = 6$

Exacta: $P = \frac{6!}{(6-2)!} = 30$

Trifecta: $P = \frac{6!}{(6-3)!} = 120$

6 horses in order: $P = \frac{6!}{(6-6)!} = 720$

# COMBINATIONS:
**Possible outcomes *irrespective* of order**

Assume $N$ = 6 horses

**First place:** $C$ = 6 possible winners

**Places 1-2:** $C$ = 15

| | | | | |
|---|---|---|---|---|
| 1-2 | 2-3 | 3-4 | 4-5 | 5-6 |
| 1-3 | 2-4 | 3-5 | 4-6 | |
| 1-4 | 2-5 | 3-6 | | |
| 1-5 | 2-6 | | | |
| 1-6 | | | | |

**Places 1-2-3:** $C$ = 20

| | | |
|---|---|---|
| 1-2-3 | 1-4-5 | 2-4-6 |
| 1-2-4 | 1-4-6 | 2-5-6 |
| 1-2-5 | 1-5-6 | 3-4-5 |
| 1-2-6 | 2-3-4 | 3-4-6 |
| 1-3-4 | 2-3-5 | 3-5-6 |
| 1-3-5 | 2-3-6 | 4-5-6 |
| 1-3-6 | 2-4-5 | |

*C*: Number of possible **COMBINATIONS**

*N*: Number of trials, events, participants, etc

*x*: Number of outcomes, order does not matter

$$C = \frac{N!}{(N-x)!\; x!} = \binom{N}{x}$$

Possible winners: $C = \frac{6!}{(6-1)!\; 1!} = 6$

Possible top-2 finishers: $C = \frac{6!}{(6-2)!\; 2!} = 15$

Possible top-3 finishers: $C = \frac{6!}{(6-3)!\; 3!} = 20$

Possible top-6 finishers: $C = \frac{6!}{(6-6)!\; 6!} = 1$

**5 of 69 numbers:**

$$C = \frac{69!}{(69-5)!\,5!} = \frac{69 \times 68 \times 67 \times 66 \times 65}{5 \times 4 \times 3 \times 2 \times 1} = 11,238,513$$

**Prize: $1,000,000**

**5 of 69 numbers:**

$$C = \frac{69!}{(69-5)!\,5!} = \frac{69 \times 68 \times 67 \times 66 \times 65}{5 \times 4 \times 3 \times 2 \times 1} = 11,238,513$$

**1 of 26 numbers:**

**26 x 11,238,513 = 292,201,338**

# COMBINATIONS: Same coin tossed 3 times

# COMBINATIONS: Same coin tossed 3 times

**All 3 tosses are heads:** $$C = \frac{3!}{(3-3)!\ 3!} = 1$$

# COMBINATIONS: Same coin tossed 3 times

**2 of 3 tosses are heads:** $$C = \frac{3!}{(3-2)!\;2!} = 3$$

# COMBINATIONS: Same coin tossed 3 times

**1 of 3 tosses are heads:**

$$C = \frac{3!}{(3-1)!\ 1!} = 3$$

# COMBINATIONS: Same coin tossed 3 times

**0 of 3 tosses are heads:** $C = \frac{3!}{(3-0)!\ 0!} = 1$

# PROBABILITIES: Same coin tossed 3 times

$$P_B = \frac{N!}{(N-x)!\; x!}\; p^x (1-p)^{N-x}$$

Heads: $p = 1/2$;
Tails: $1 - p = 1/2$

$P_B$: BINOMIAL DISTRIBUTION

**3 of 3 tosses are heads:** $P_B(x=3) = 1 \times \; (1/2)^3 (1/2)^{3-3} = \frac{1}{8}$

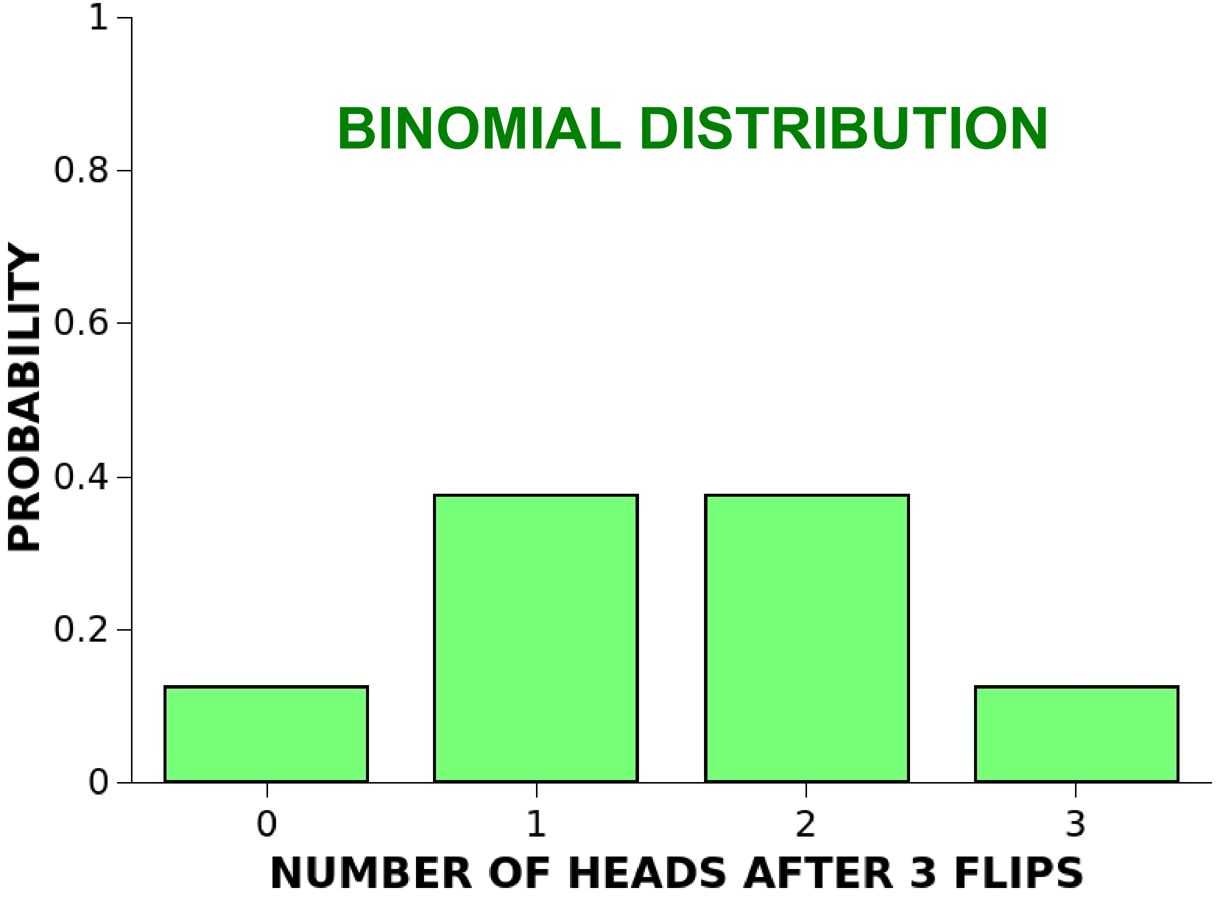
**2 of 3 tosses are heads:** $P_B(x=2) = 3 \times \; (1/2)^2 (1/2)^{3-2} = \frac{3}{8}$

**1 of 3 tosses are heads:** $P_B(x=1) = 3 \times \; (1/2)^1 (1/2)^{3-1} = \frac{3}{8}$

**0 of 3 tosses are heads:** $P_B(x=0) = 1 \times \; (1/2)^0 (1/2)^{3-0} = \frac{1}{8}$

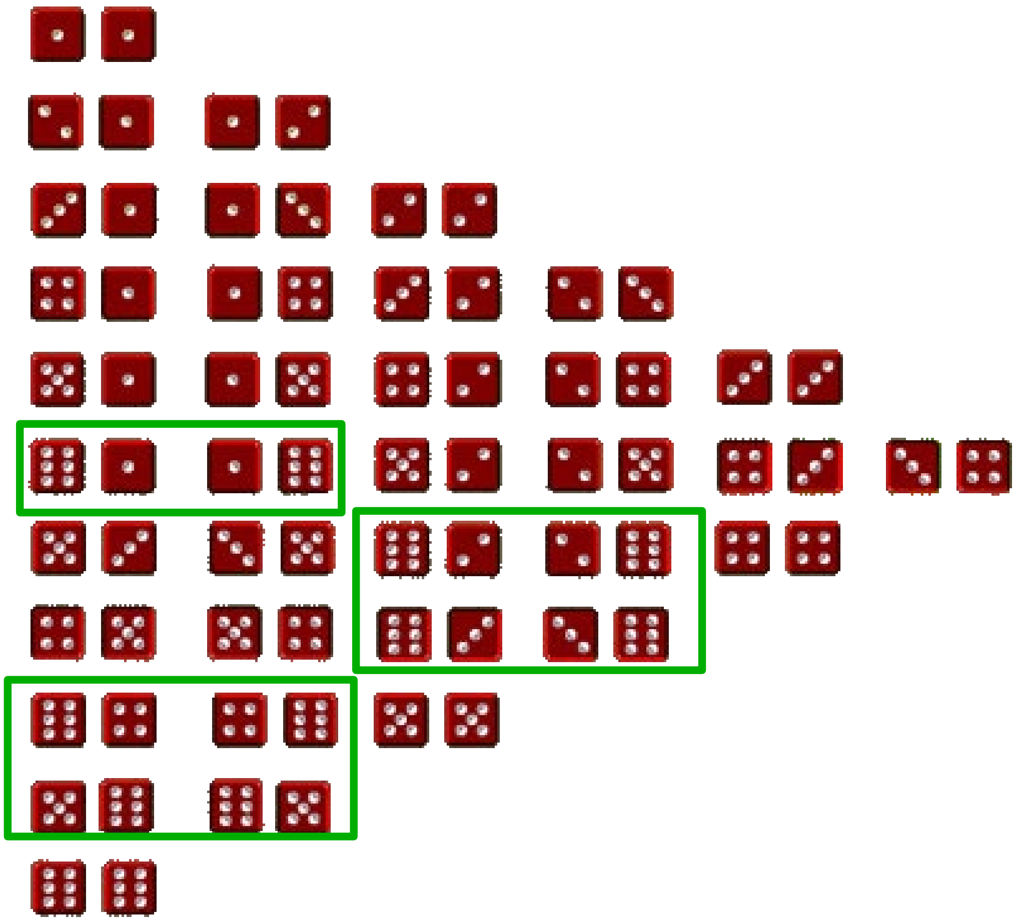
Probabilities sum to 1

# PROBABILITY DISTRIBUTION:
## Number of HEADS occurring on 3 consecutive coin flips



**PROBABILITY** that exactly *x*=1 **SIX** appears in *N*=2 rolls of the die [or one roll of two dice]:

$$P_B = \frac{2!}{1!(2-1)!} \times \left(\frac{1}{6}\right)^1 \left(1-\frac{1}{6}\right)^{2-1} = \frac{10}{36}$$

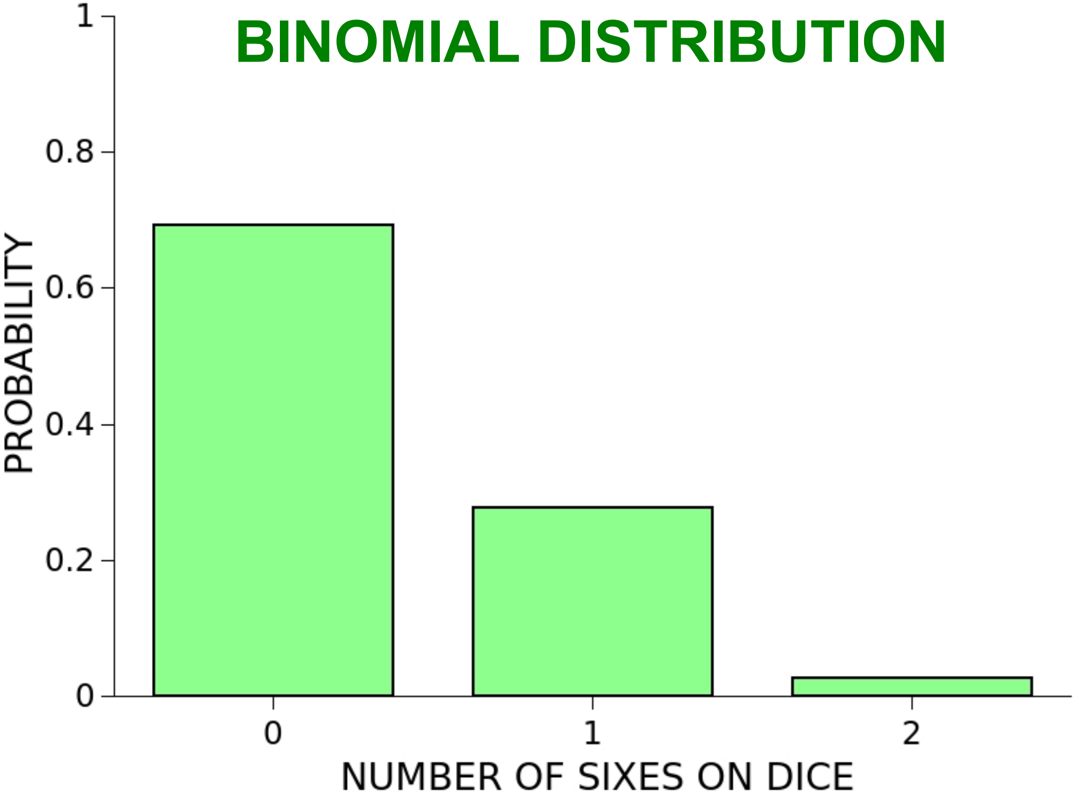


# PROBABILITY DISTRIBUTION:
## SIX appearing on pair of dice

Probability of zero SIXES: 25/36

Probability of one SIX: 10/36

Probability of two SIXES: 1/36



# Toss same coin tossed *N* = 10 times

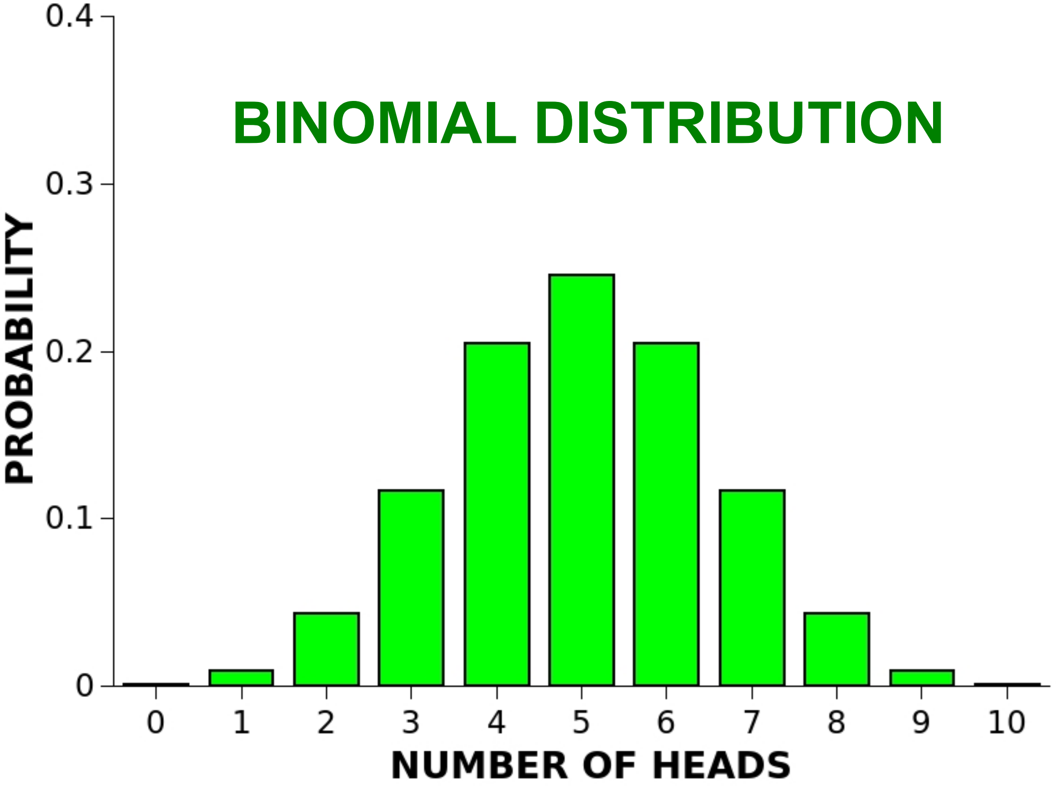
## *x*: Number of times HEADS appears

$$P_B = \frac{N!}{(N-x)!\; x!}\; p^x (1-p)^{N-x}$$

Heads: $p$ = 1/2;
Tails: 1 – $p$ = 1/2

# PROBABILITY DISTRIBUTION:

## Number of HEADS occurring on 10 consecutive coin

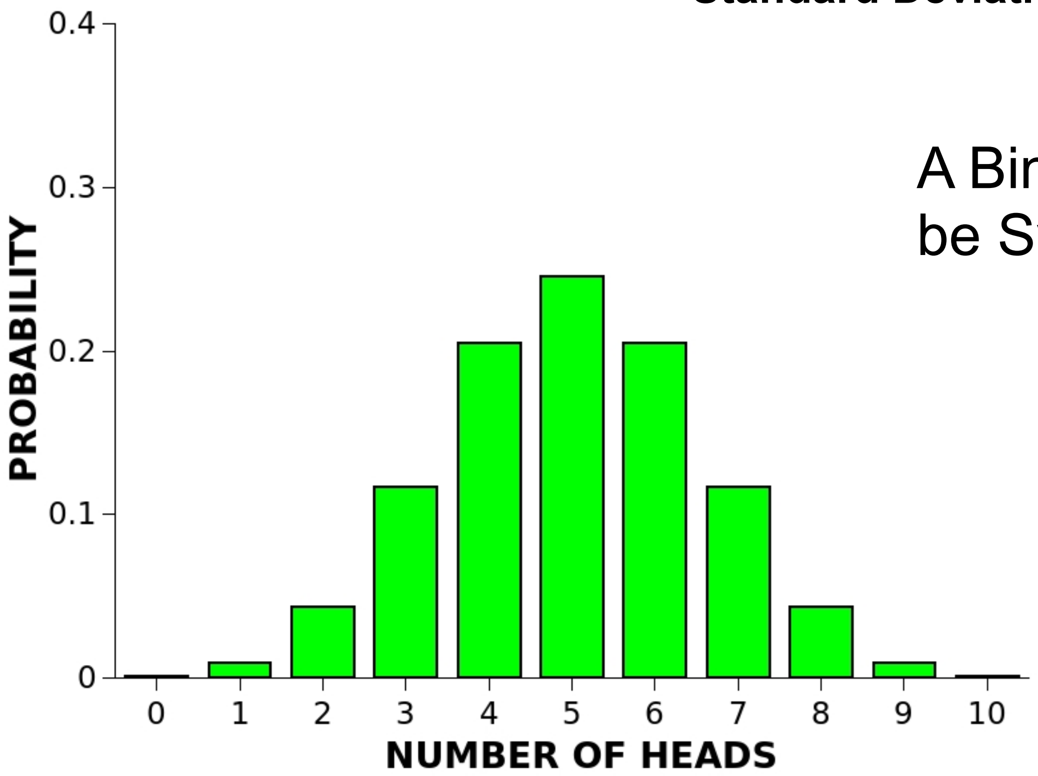


# BINOMIAL DISTRIBUTION

**Mean:** $Np = 5$

**Variance:** $\sigma^2 = Np(1-p) = 2.5$

**Standard Deviation:** $\sqrt{\sigma} = \sqrt{Np(1-p)} = 1.58$




A Binomial Distribution may be Symmetric or Asymmetric

# POISSON DISTRIBUTION

An approximation to the Binomial distribution

Probability *p* gets small

Large number trials: *N* is big

Typically: Counting *x* events occurring in a time interval

Events individually distinguishable; uncorrelated

Mean rate: l = *Np*

Standard deviation: $\sigma = \sqrt{\lambda}$

$$P_P = \frac{\lambda^x}{x!} e^{-\lambda}$$

# EXAMPLE: NUCLEAR DECAY

Half-life: Multiple years $\longrightarrow$ Decay probability *p* very small

Number of nucleii *N* very large

Mean rate: l = *Np;*
...but *N* and *p* are likely unknown!

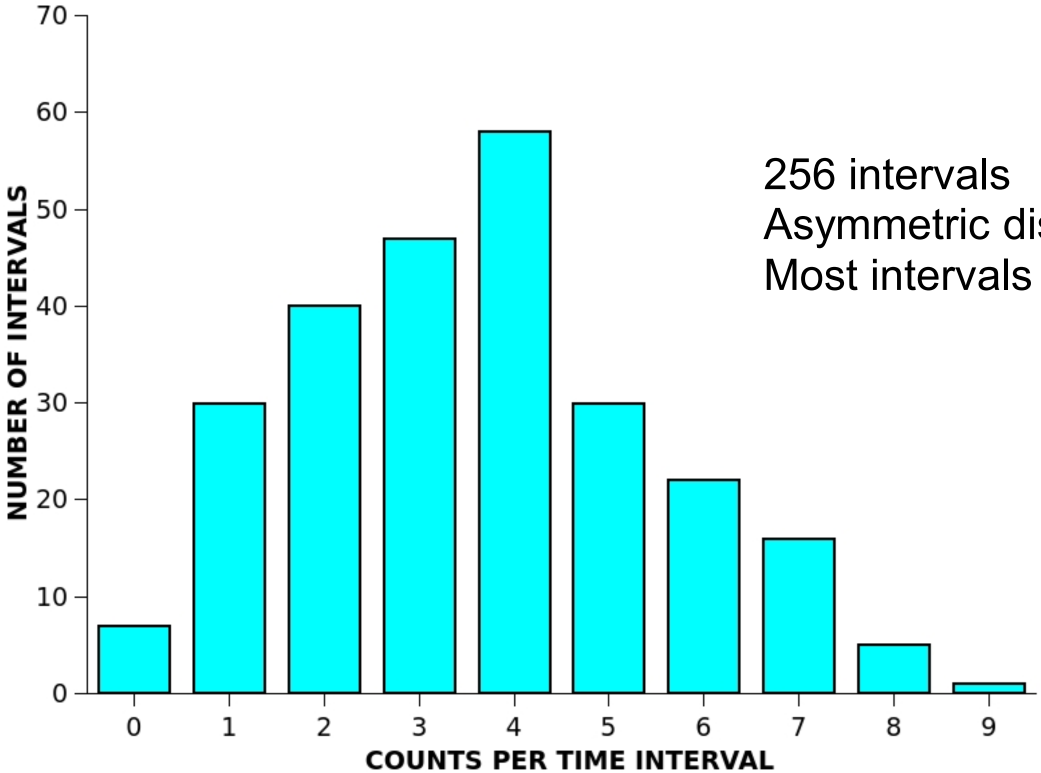
l = Total events counted / Total observation time

$$P_P = \frac{\lambda^x}{x!}e^{-\lambda}$$

# EXAMPLE: NUCLEAR DECAY

Count number of radioactive decays $x$ in a series of intervals of duration t

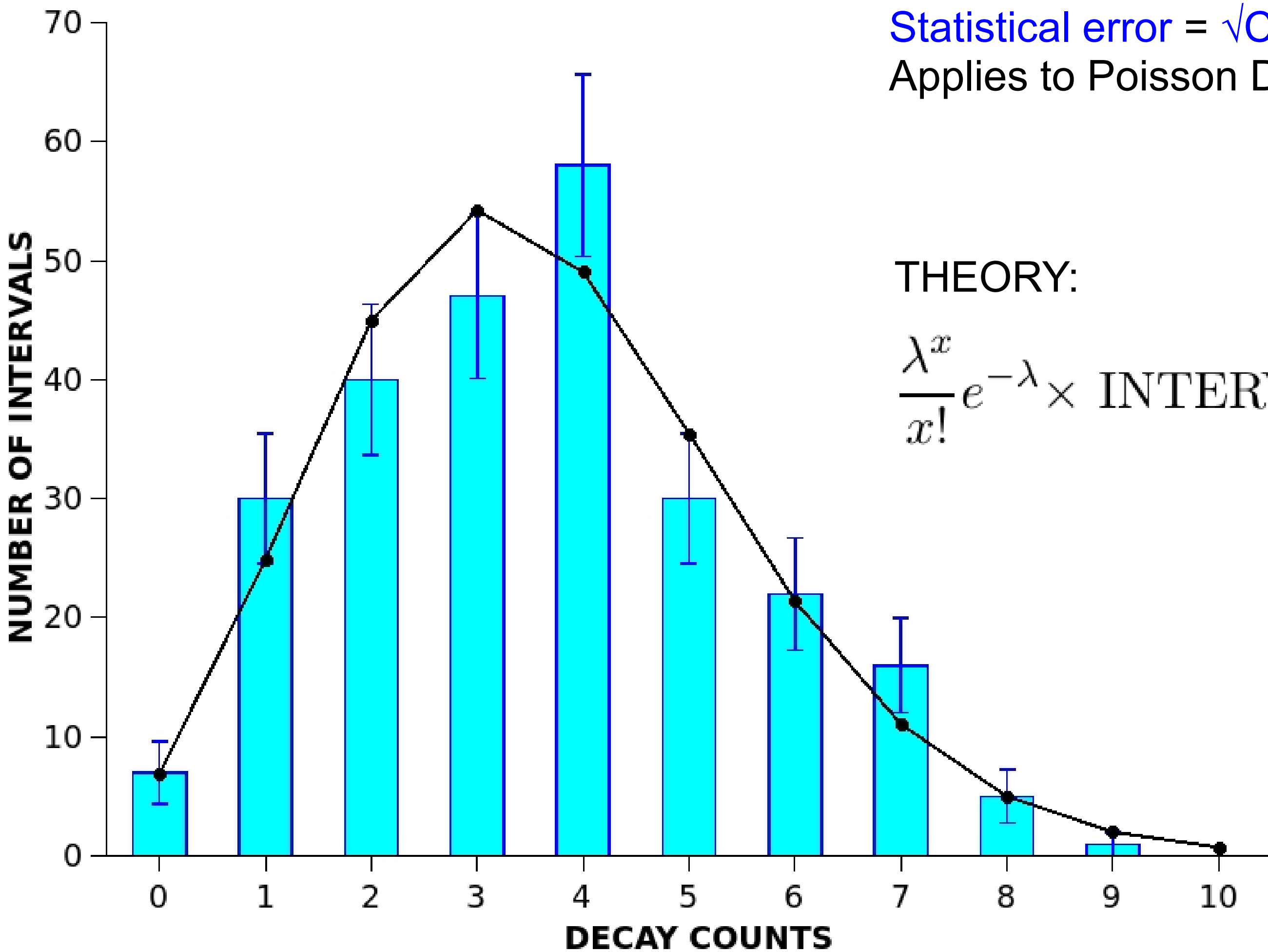
Plot on a histogram:




256 intervals
Asymmetric distribution
Most intervals count $x$ = 4 decays

# EXAMPLE: NUCLEAR DECAY

## Comparing experiment with theory



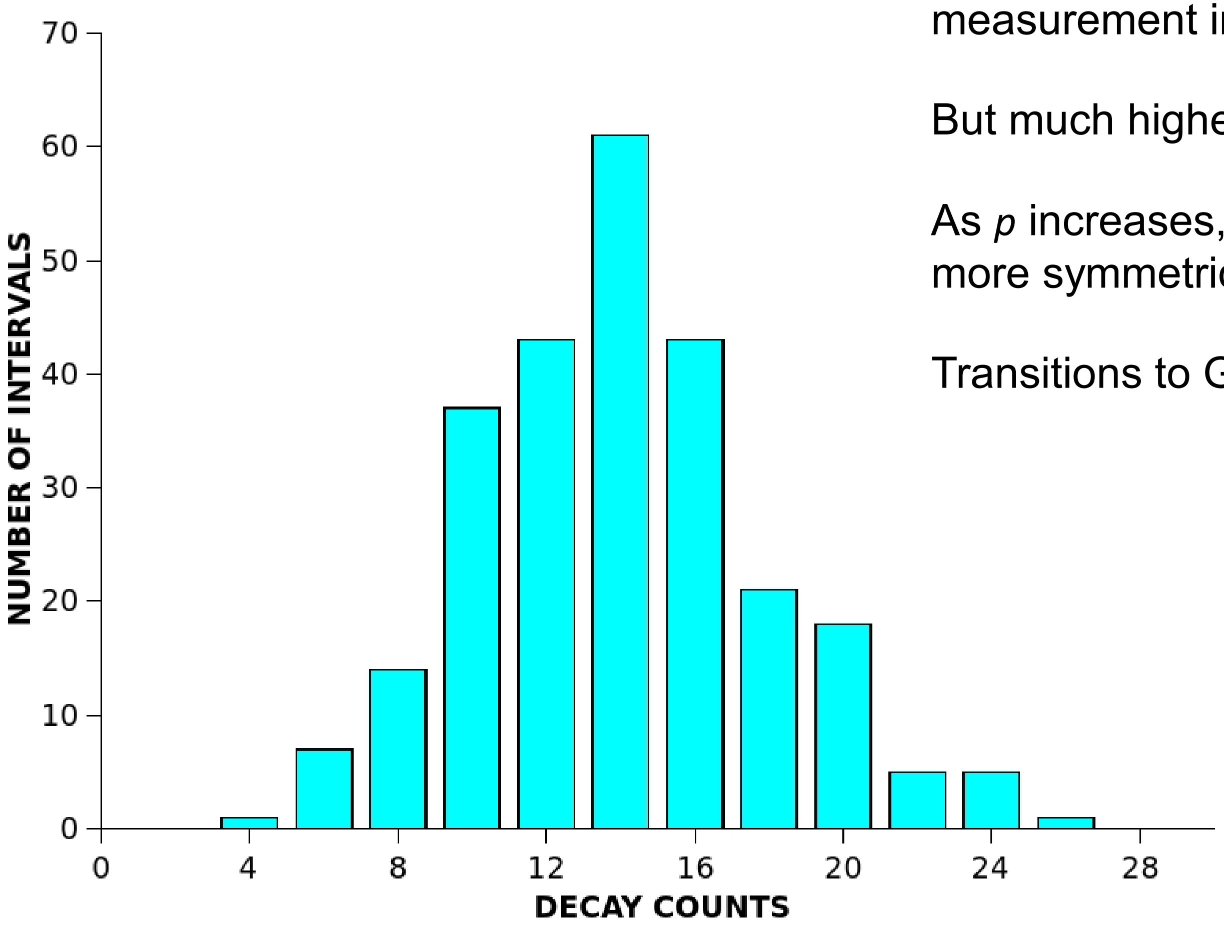

Statistical error = $\sqrt{\text{Counts}}$

Applies to Poisson Distribution only!

THEORY:

$$\frac{\lambda^x}{x!} e^{-\lambda} \times \text{INTERVALS}$$

# EXAMPLE: NUCLEAR DECAY




Experiment repeated with same number of measurement intervals

But much higher count rate

As *p* increases, distribution becomes more symmetric

Transitions to Gaussian

# PROBABILITY DISTRIBUTION FUNCTIONS

- **GAUSSIAN:** Random data, experimental parameters uncertain
- **POISSON:** Number of counts in a specified time interval
- **BINOMIAL:** Small number of possible outcomes (eg. heads or tails)

# GAUSSIAN DISTRIBUTION aka "The Bell Curve"

An approximation to the Binomial distribution

Number of trials *N* gets large

$Np >> 1$

Most experimental distributions are Gaussian

Most probable result is the **AVERAGE** result

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$

$\overline{x}$ : Average or mean of the data

$\sigma$ : Standard deviation of the data

# GAUSSIAN DISTRIBUTION aka “The Bell Curve”

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$

Peak of curve: $x = \overline{x}$ $\qquad \overline{x} = \frac{1}{N}\sum_i x_i$

$$\sigma^2 = \frac{1}{N-1}\sum_i^N (x_i - \overline{x})^2$$

When we average a set of data, the **implicit assumption** is a Gaussian Distribution

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$

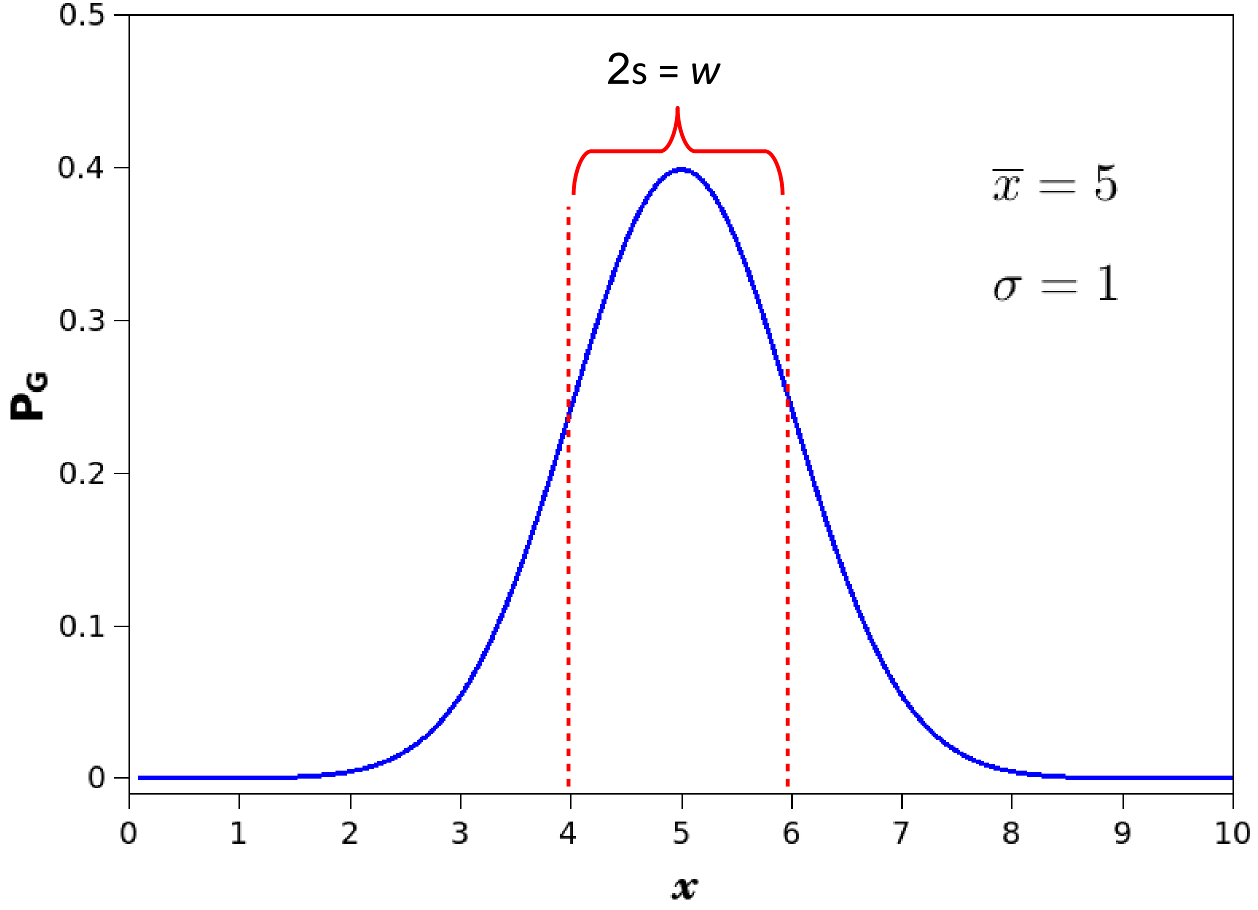
**CAUTION: Sometimes written with *w***

$$P_G = \frac{1}{w}\sqrt{\frac{2}{\pi}} \exp\left[-2\left(\frac{x-\overline{x}}{w}\right)^2\right]$$

$$w = 2\sigma$$

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$

There is a 68% chance that a measurement will lie within $\overline{x} \pm \sigma$



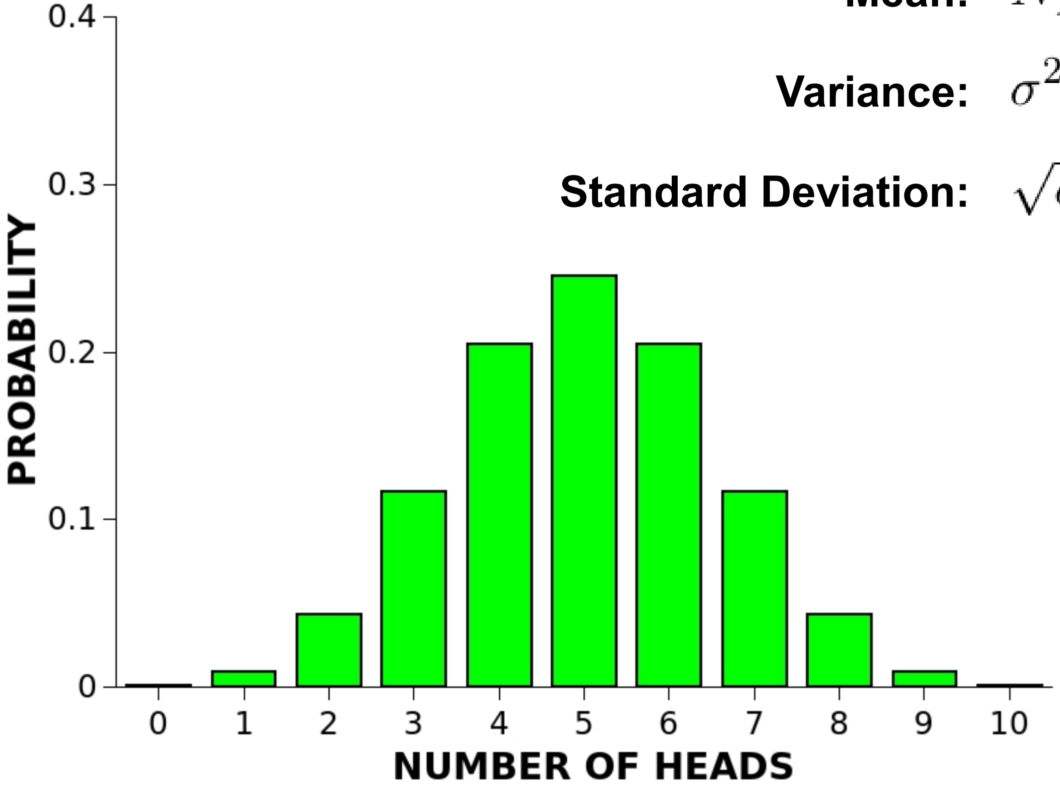
# Number of HEADS occurring on 10 consecutive coin flips

## *BINOMIAL DISTRIBUTION*

**Mean:** $Np = 5$

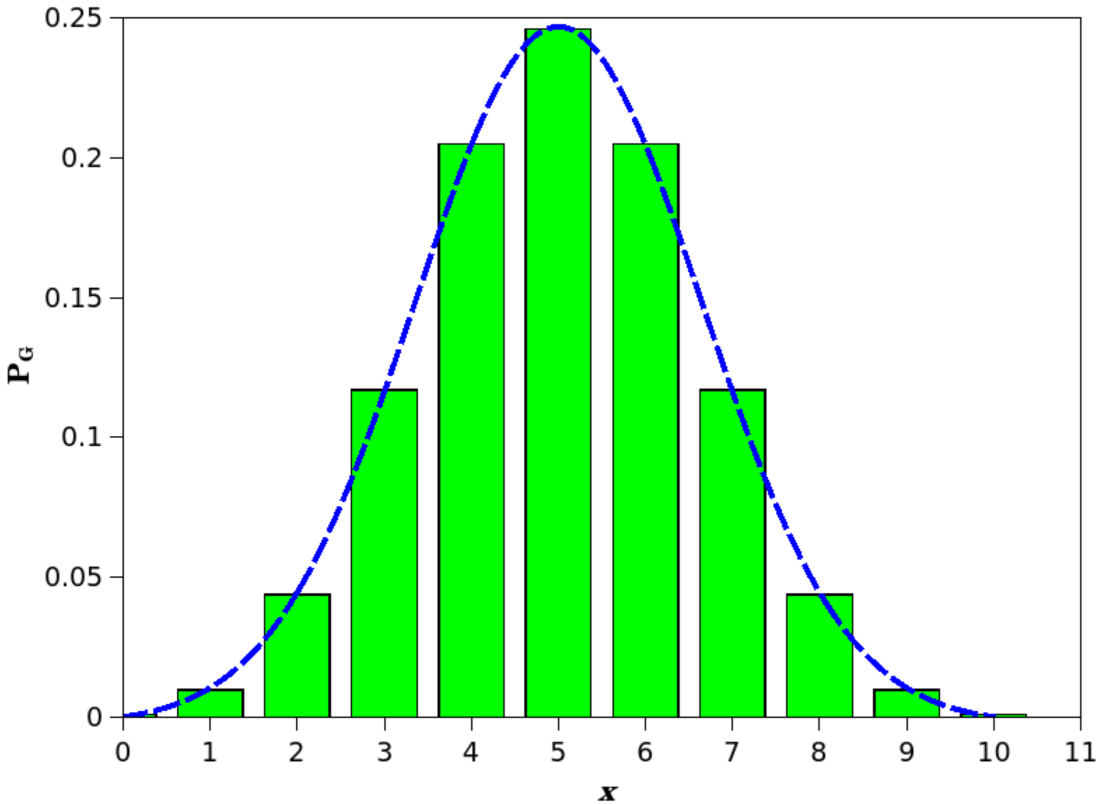
**Variance:** $\sigma^2 = Np(1-p) = 2.5$

**Standard Deviation:** $\sqrt{\sigma} = \sqrt{Np(1-p)} = 1.58$



# Fitting with a Gaussian

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$



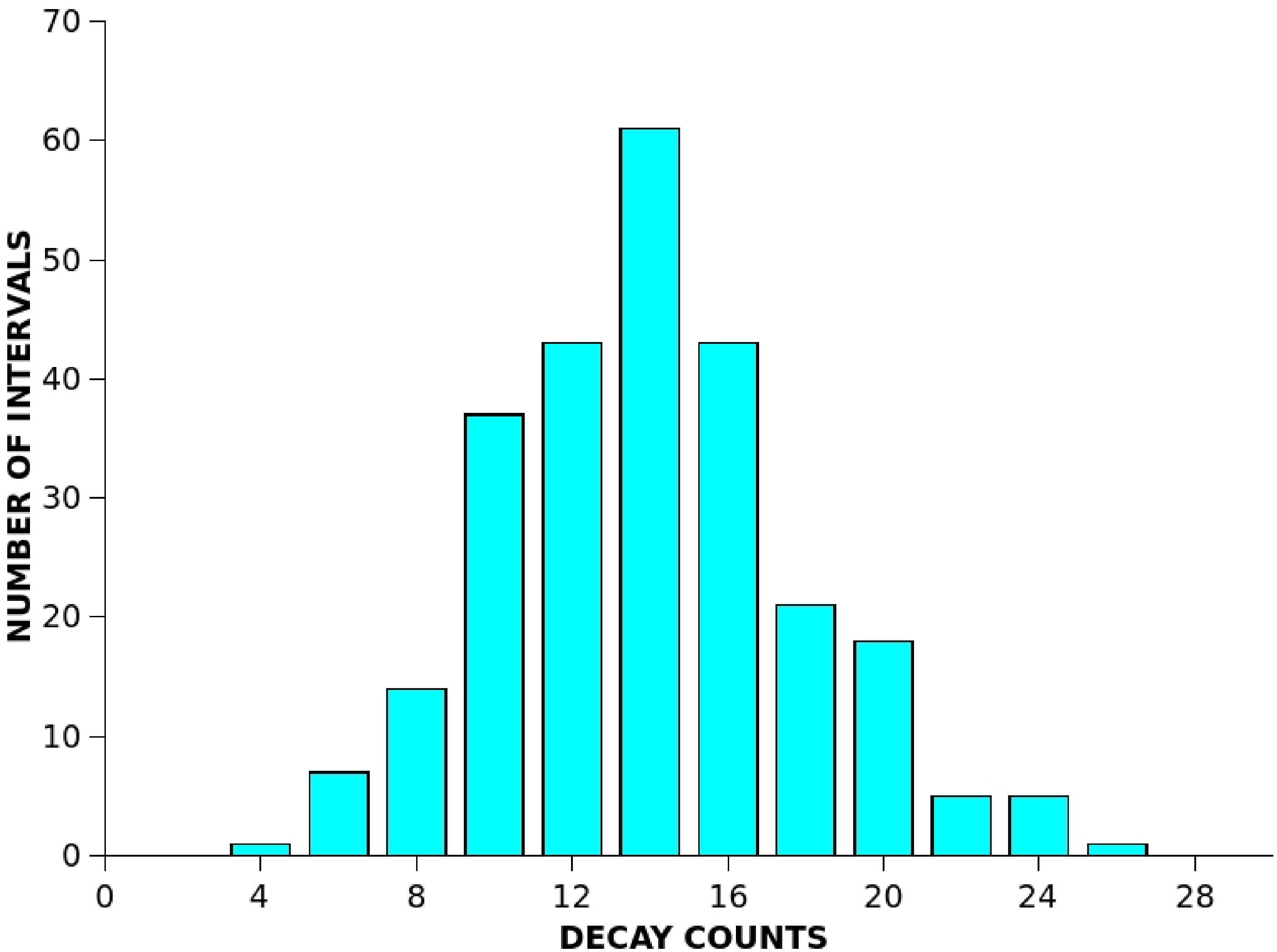

$$\overline{x} = \frac{1}{N}\sum_{i}^{N} x_i$$

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i}(x_i - \overline{x})^2}$$

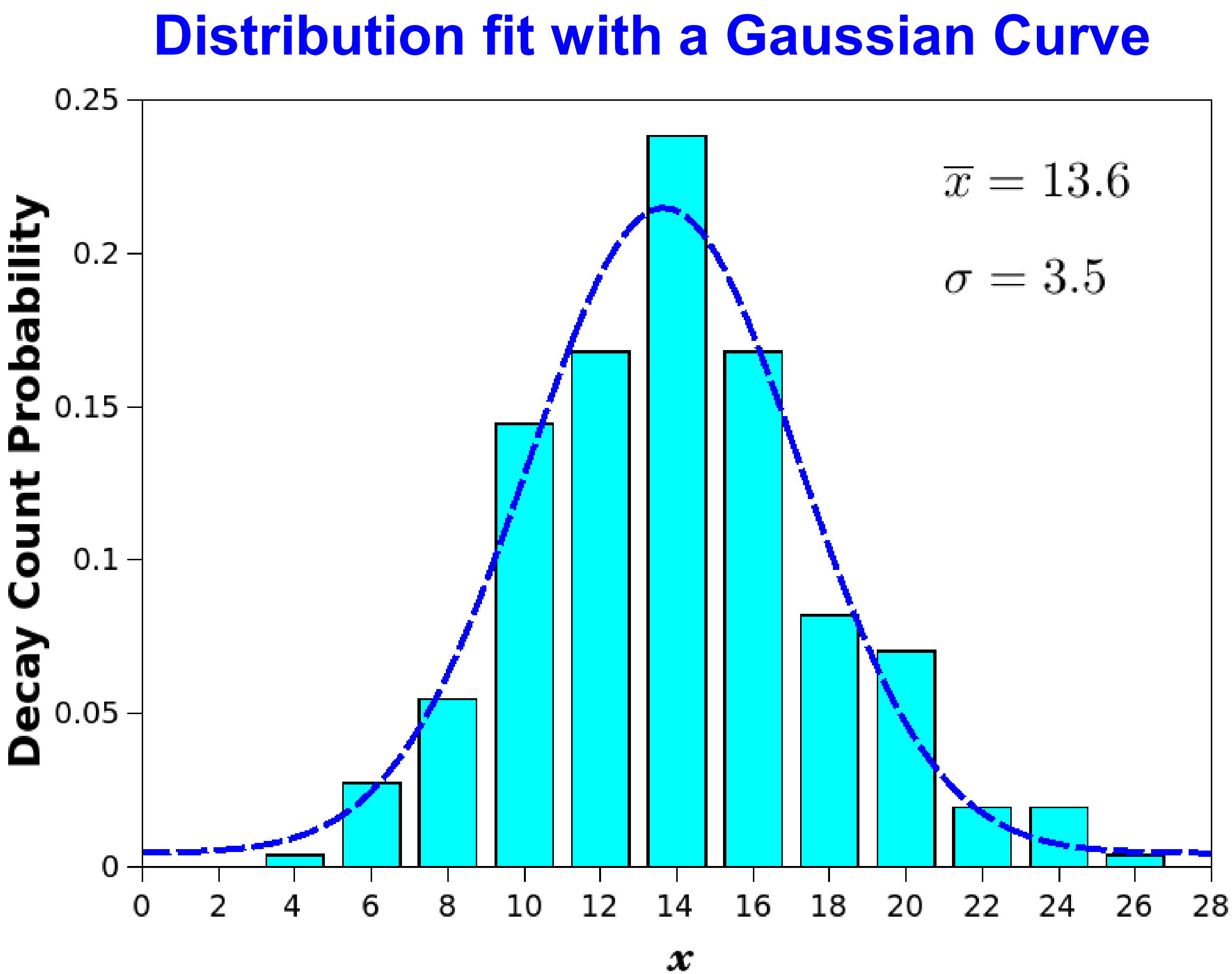
$$\overline{x} = 5$$

$$\sigma = 1.64$$

# Experimental Radioactive Decay Data

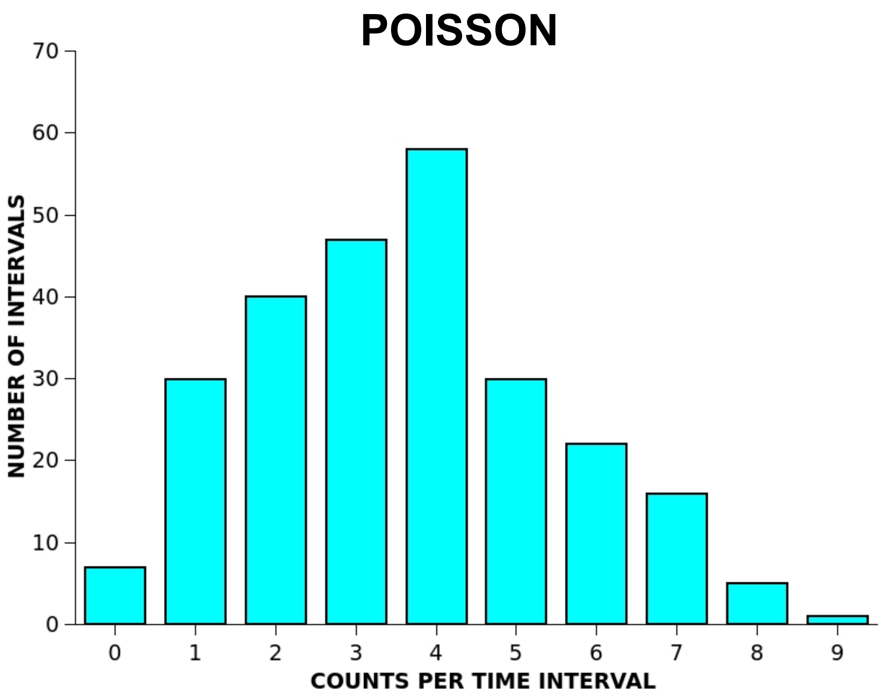


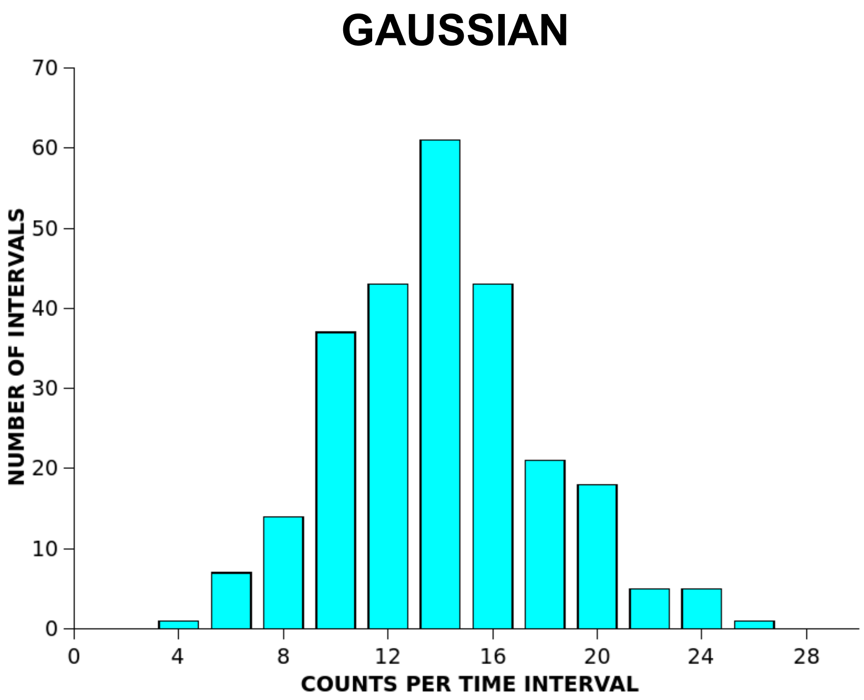
# Experimental Radioactive Decay Data



***Recall that:***

**Poisson transitions to Gaussian as data count rate increases**

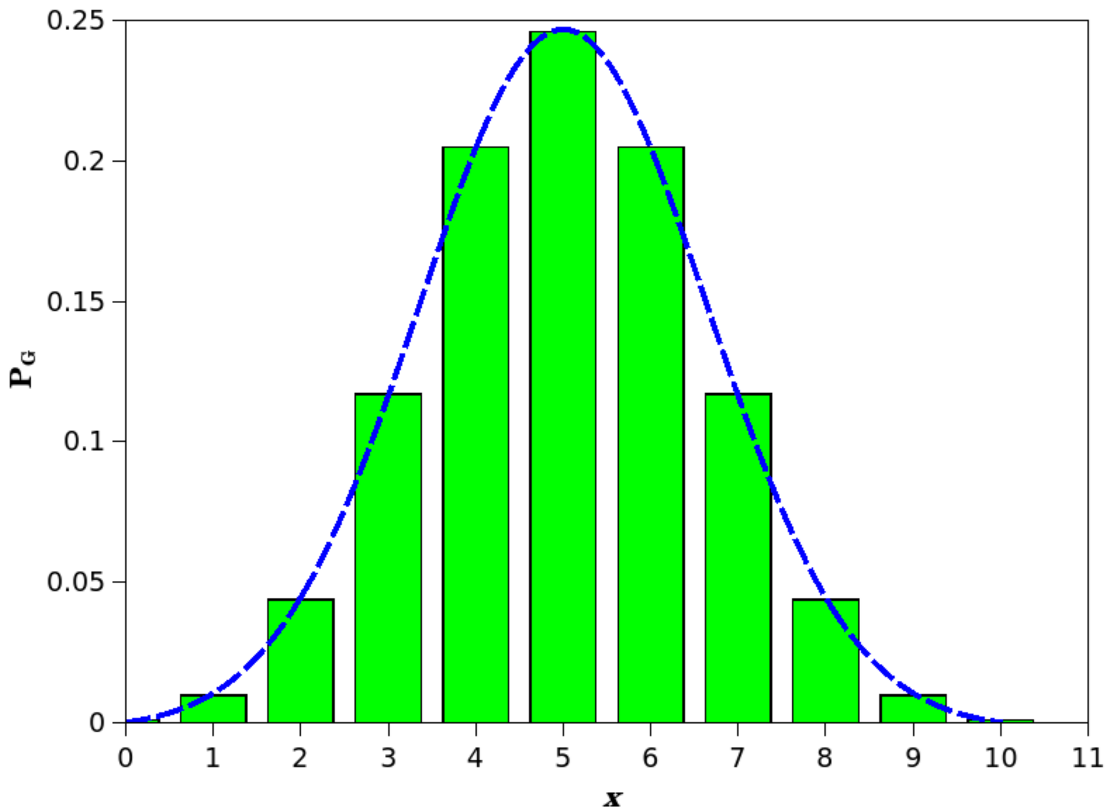


# Uncertainty of the *Mean Value*: $\overline{x} \pm ?$

- Gaussian distribution; *N* data points
- Uncertainty of distribution: s
- Uncertainty in ***Mean*** decreases with *N*

$$\overline{x} \pm \frac{\sigma}{\sqrt{N}}$$




*N:* 10 coin flips

*x:* Number of heads occurring

$$\sigma = 1.64$$

$$\overline{x} = 5 \pm \frac{\sigma}{\sqrt{N}} = 5 \pm 0.52$$

## Implications of increasing *N*

$$\overline{x} \pm \frac{\sigma}{\sqrt{N}}$$

Assumes all data in distribution has same uncertainty

As $N \longrightarrow \infty$, accuracy becomes perfect i.e. no error!

Acquiring huge amount of data may not be possible

Experiment may drift with time: Systematic error

Very difficult to eliminate all systematic errors

# Comparing Distribution Functions

**Binomial**: Probability of observing *x* in *N* trials when the probability *p* of *x* occurring is known

$$P_B = \frac{N!}{(N-x)!\; x!}\; p^x (1-p)^{N-x}$$

**Poisson**: Approximation to Binomial

Values of *x* are strictly bounded $x \geq 0$

Primary useful for low data/count rates

Standard deviation: $\sigma = \sqrt{\lambda}$

Asymmetric distributions

$$P_P = \frac{\lambda^x}{x!} e^{-\lambda}$$

**Gaussian**: Approximation to Binomial

Usually more convenient for analyzing experiments

$x < 0$ allowed

$$P_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\overline{x}}{\sigma}\right)^2\right]$$