LECTURE 4: Probability and Statistics (Part 2)

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)

What is the probability P_B of x events occurring in N trials
if the single event probability is p?

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

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

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)

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: $\lambda = Np$

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

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

EXAMPLE: NUCLEAR DECAY

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

Number of nucleii N very large

Mean rate: $\lambda = Np$;
...but N and p are likely unknown!

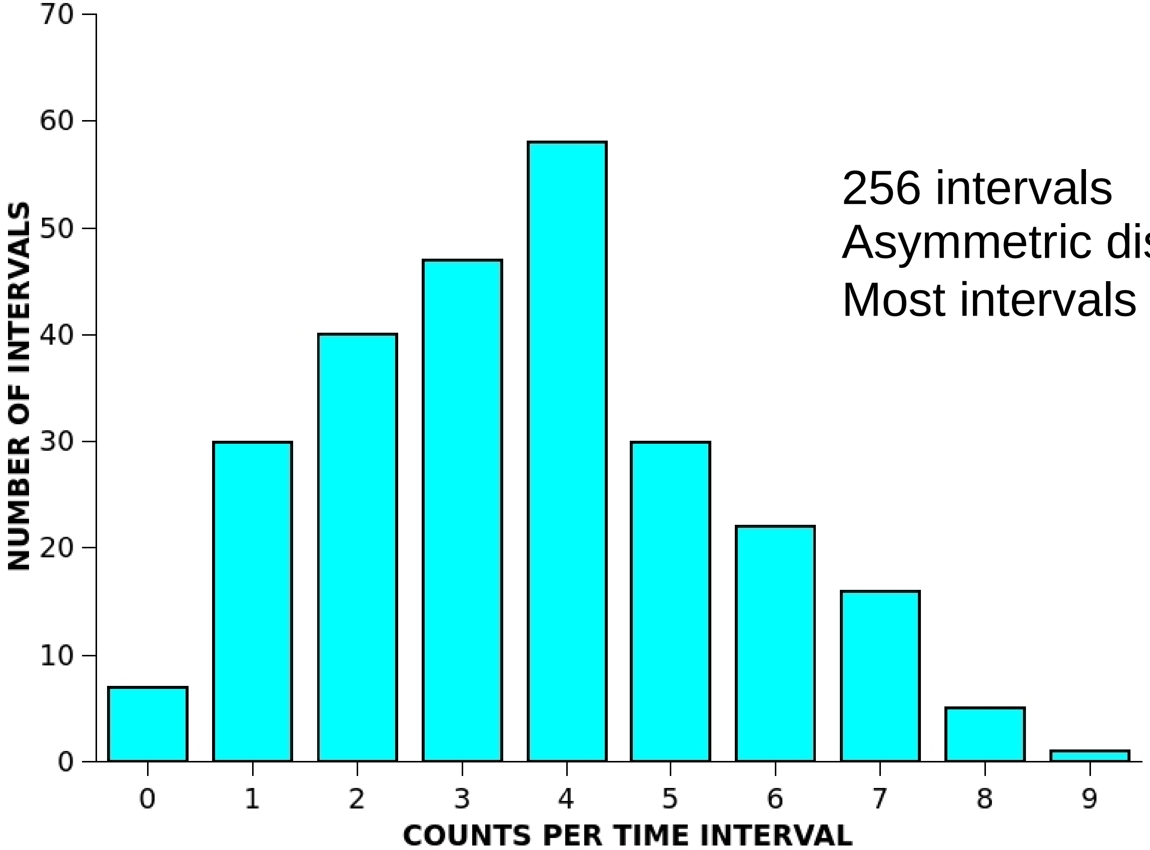
$$\lambda = \frac{\text{Total events counted}}{\text{Total observation time}}$$

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

EXAMPLE: NUCLEAR DECAY

Count number of radioactive decays x in a series of intervals of duration τ

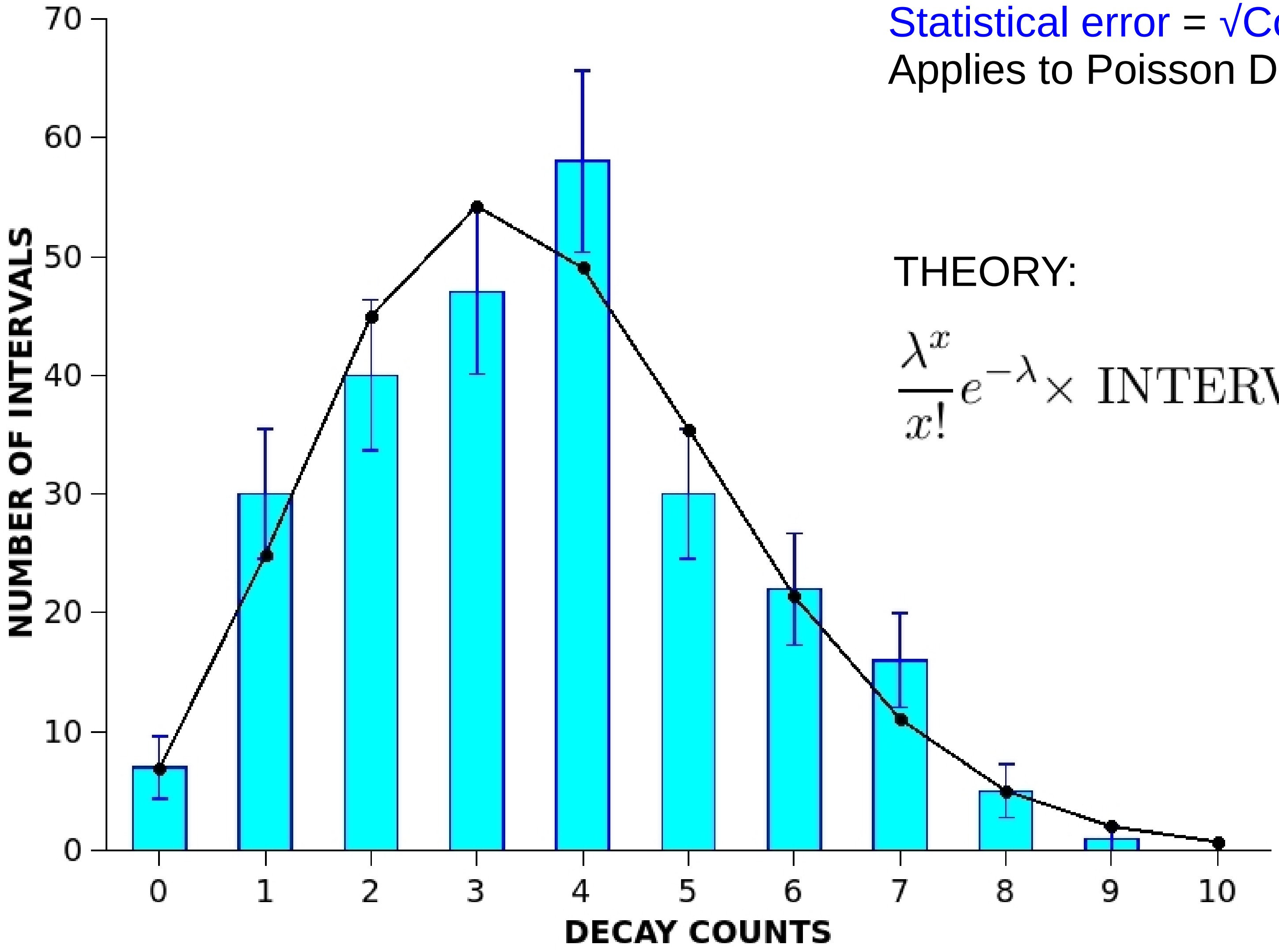
Plot on a histogram:



256 intervals
Asymmetric distribution
Most intervals count x = 4 decays

EXAMPLE: NUCLEAR DECAY

Comparing experiment with theory



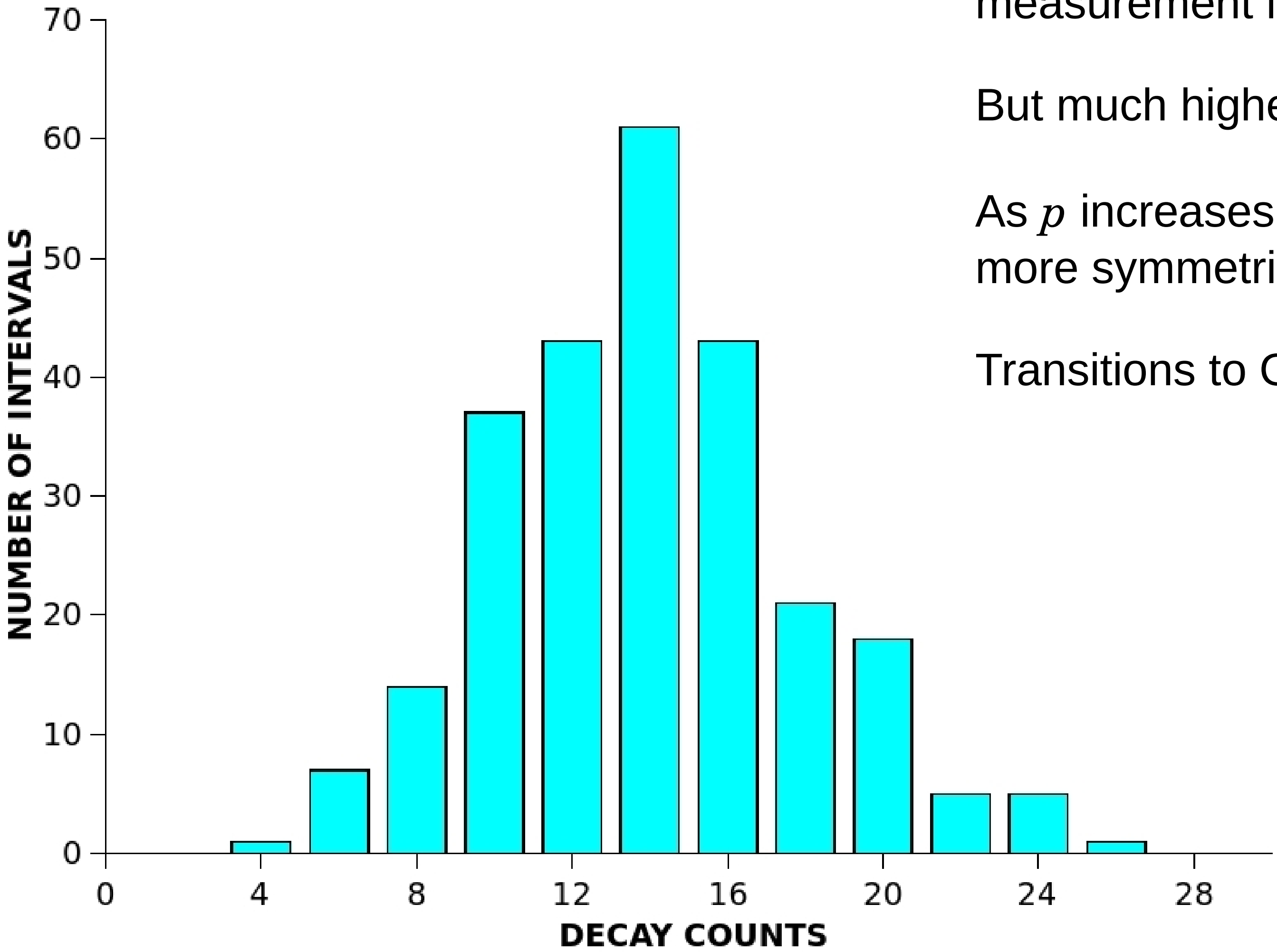
Statistical error = $\sqrt{\text{Counts}}$

Applies to Poisson Distribution only!

THEORY:

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

EXAMPLE: NUCLEAR DECAY



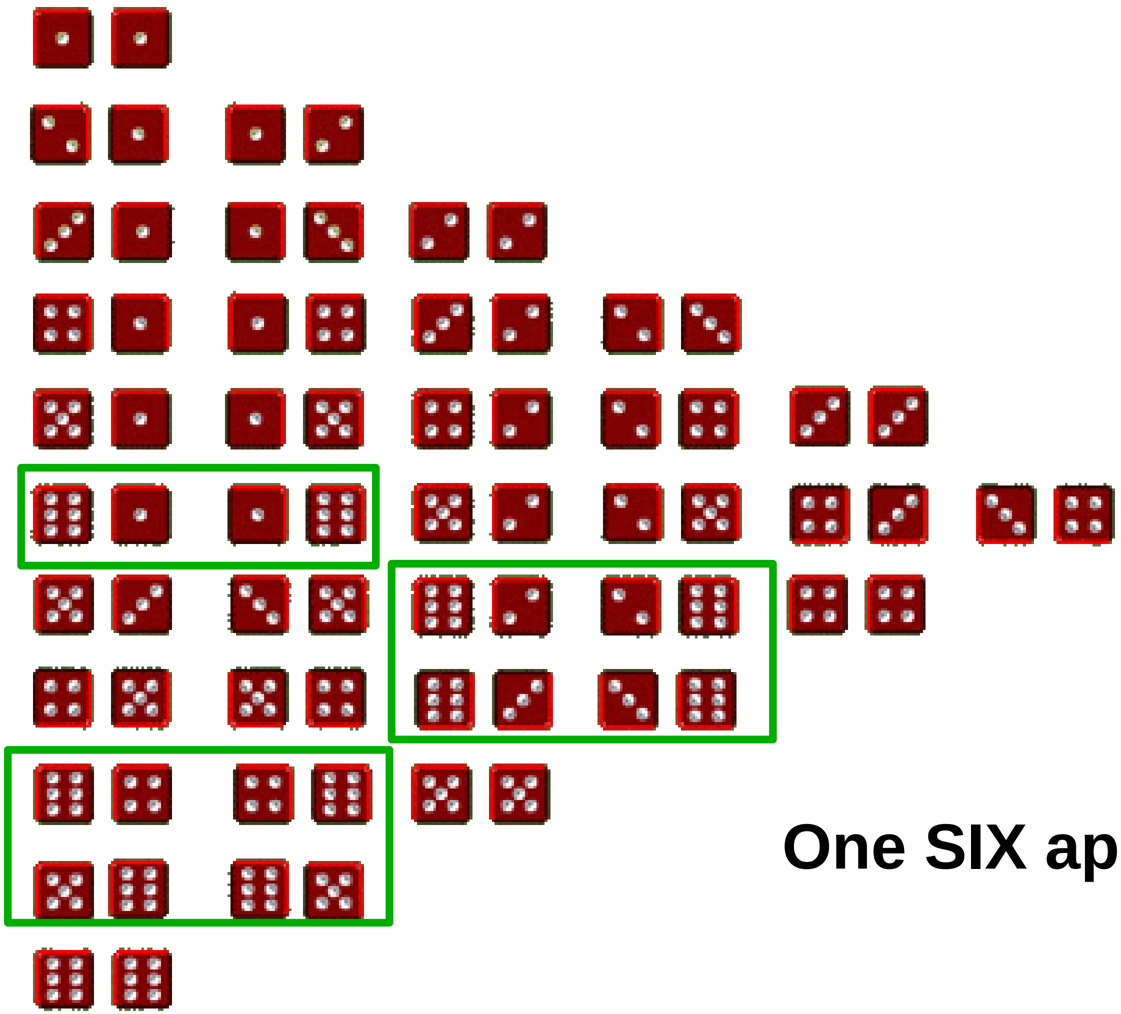
Experiment repeated with same number of measurement intervals

But much higher count rate

As p increases, distribution becomes more symmetric

Transitions to Gaussian

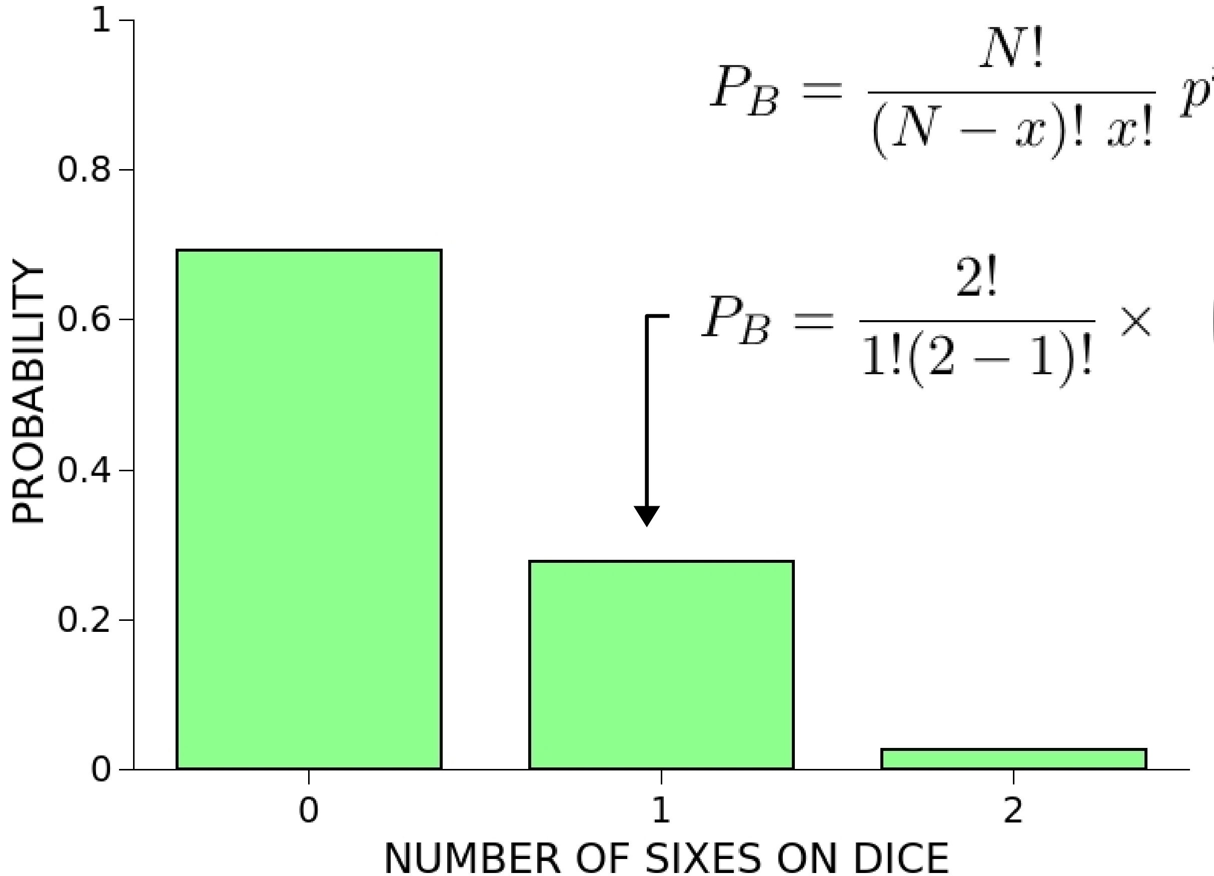
EXAMPLE: Probability of SIX appearing when dice tossed



One SIX appears 10/36

EXAMPLE: Probability of SIX appearing on dice

Exactly described by a Binomial Distribution

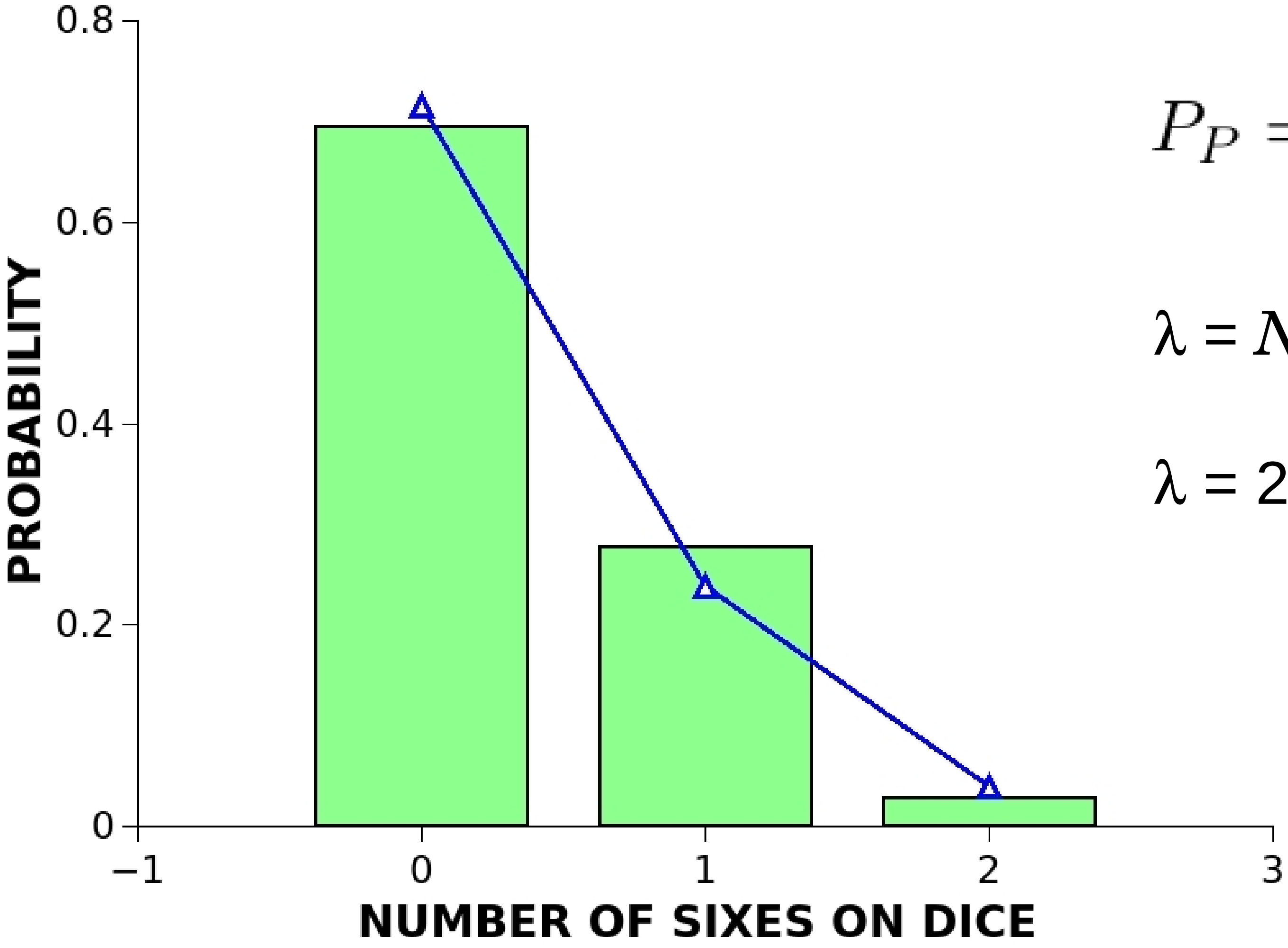


$$P_B = \frac{N!}{(N-x)!\,x!}\,p^x(1-p)^{N-x}$$

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

Can a Poisson Distribution reasonably describe the dice toss?

Poisson approximates Binomial when p small; N large



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

$\lambda = Np$ is known here

λ = 2 x 1/6 = 1/3

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

σ : 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: $x = \overline{x}$ $\quad \overline{x} = \frac{1}{N}\sum_i x_i$

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

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

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

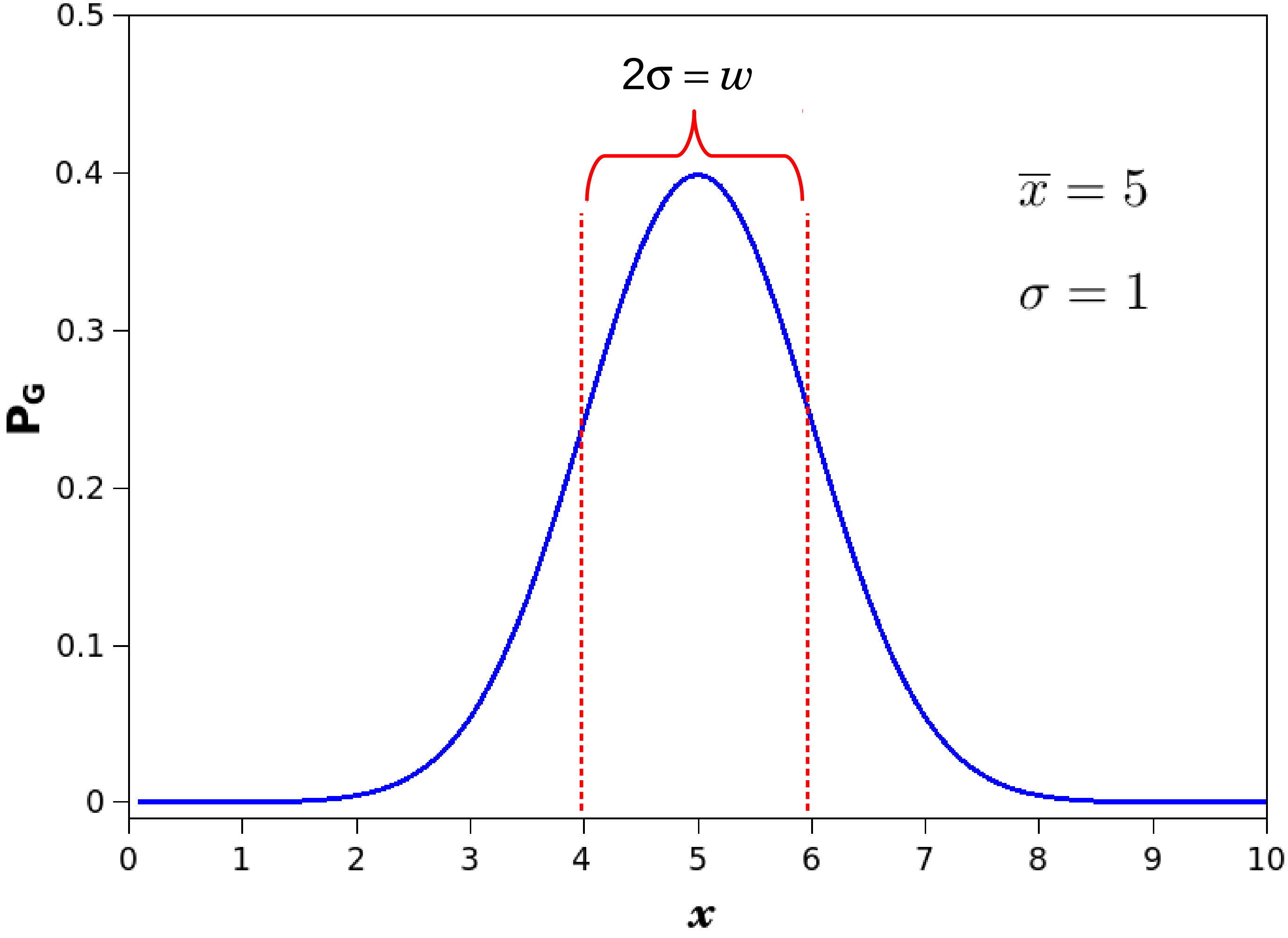
CAUTION: Sometimes written with w

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

$$w = 2\sigma$$

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

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



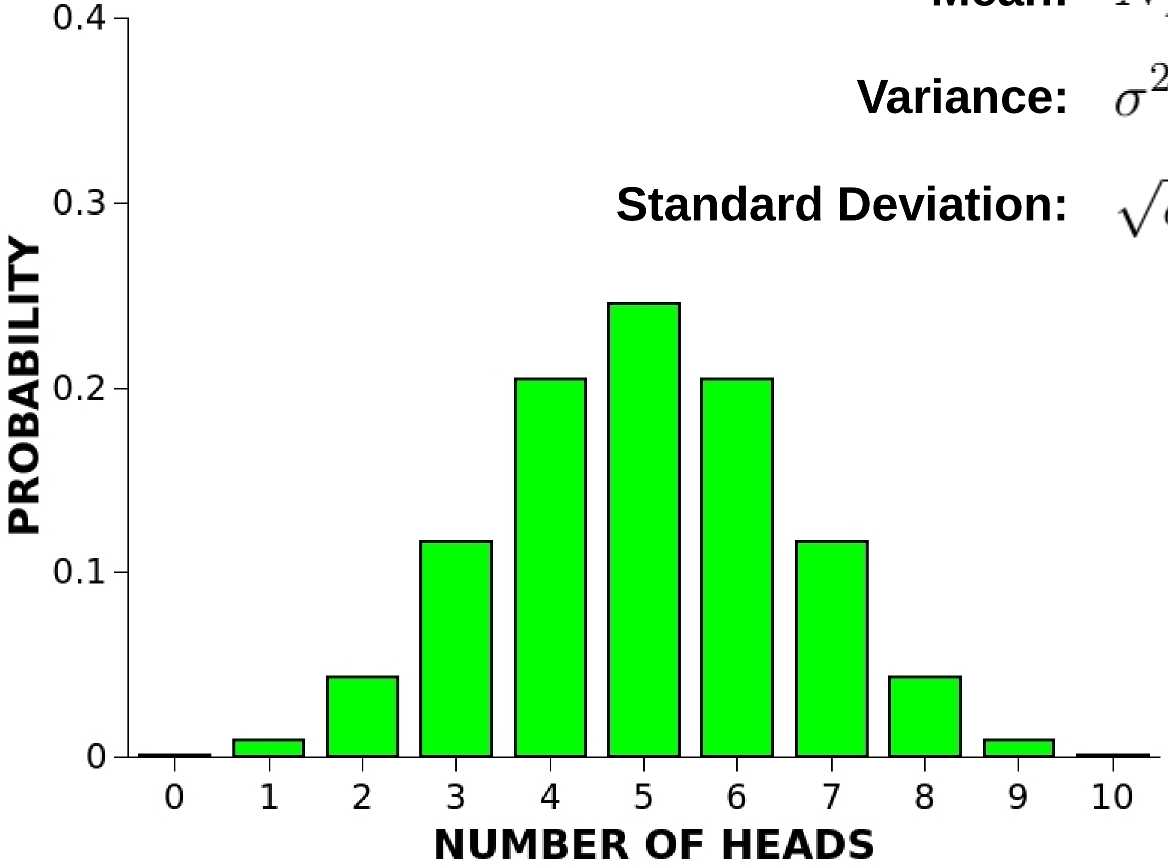
Number of HEADS occurring on 10 consecutive coin flips

BINOMIAL DISTRIBUTION

Mean: $Np = 5$

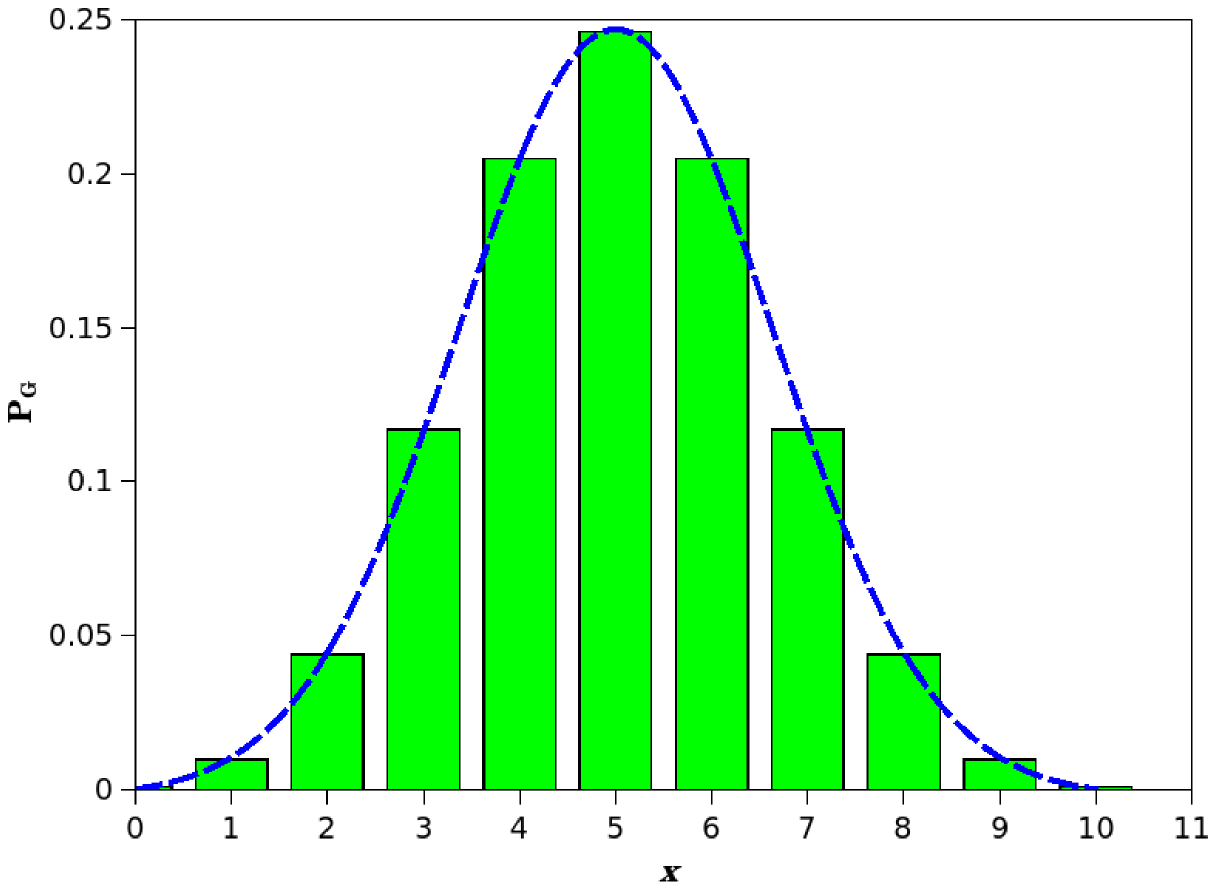
Variance: $\sigma^2 = Np(1-p) = 2.5$

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



Fitting with a Gaussian

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



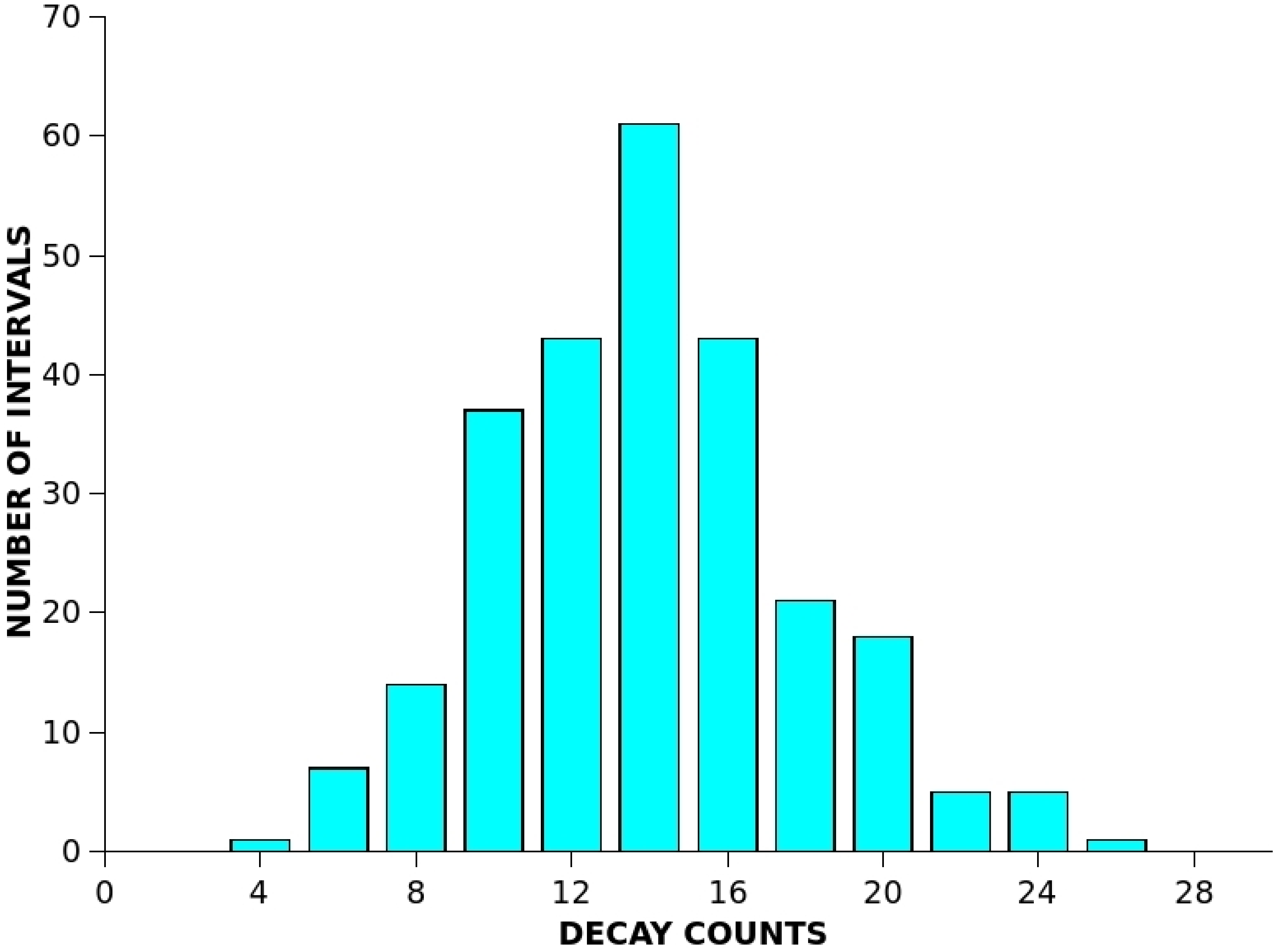
$$\overline{x} = \frac{1}{N}\sum_i^N x_i$$

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

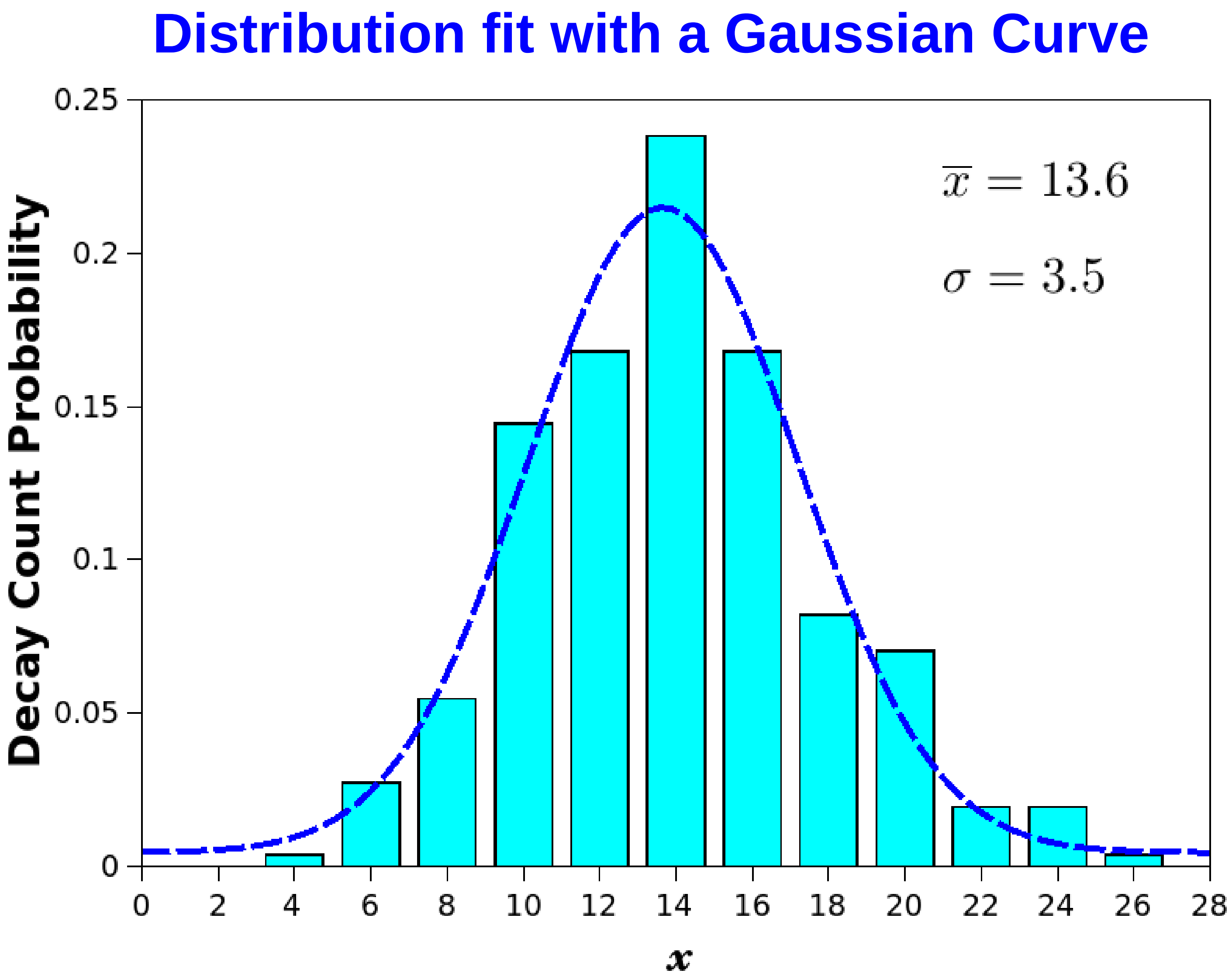
$$\overline{x} = 5$$

$$\sigma = 1.64$$

Experimental Radioactive Decay Data

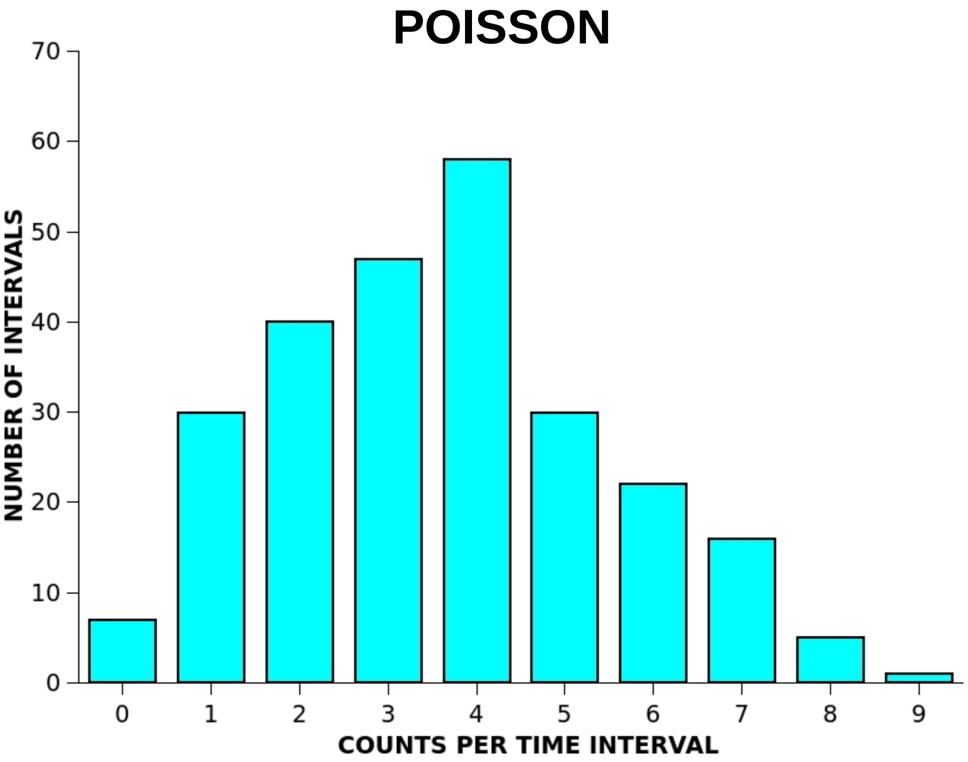


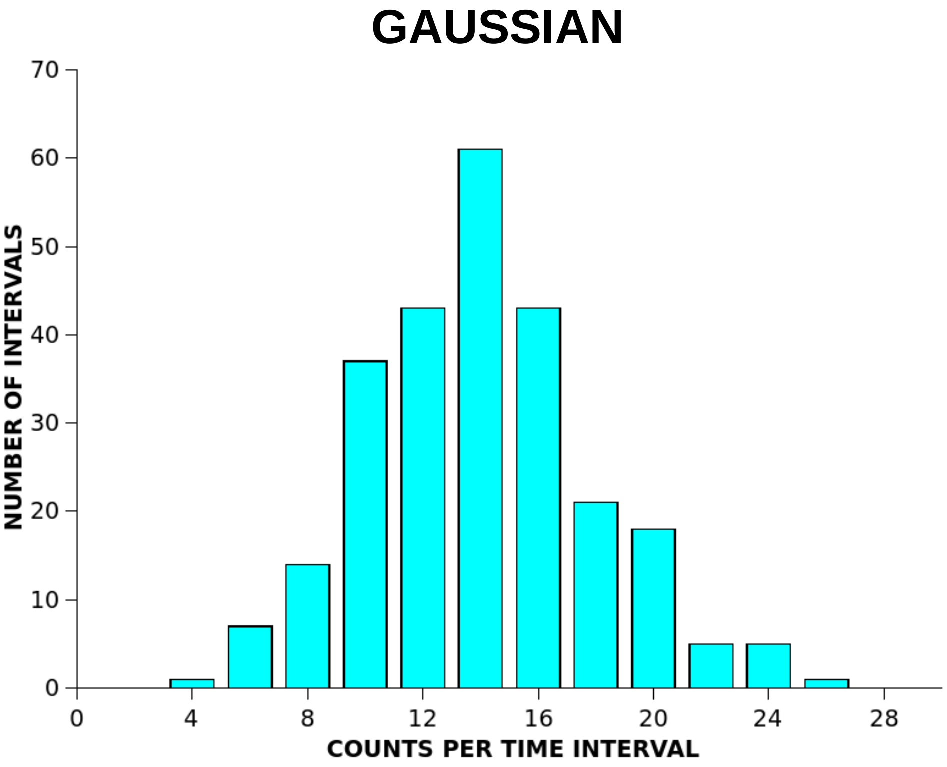
Experimental Radioactive Decay Data



Recall that:

Poisson transitions to Gaussian as data count rate increases

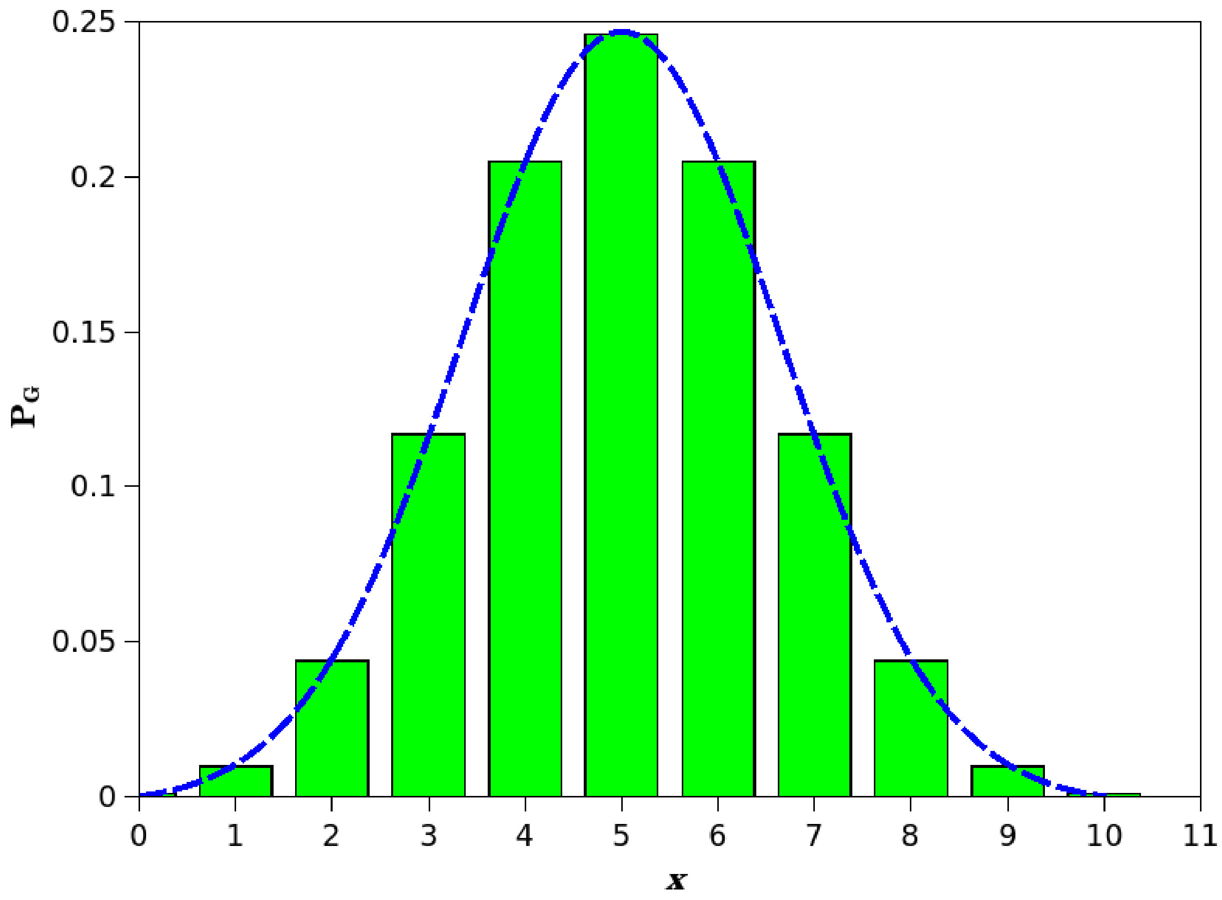




Uncertainty of the ***Mean Value***: $\overline{x} \pm ?$

- Gaussian distribution; N data points
- Uncertainty of distribution: σ
- 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 $N \rightarrow \infty$, 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

Individual data points x_i have corresponding uncertainties σ_i

Different error bars for different data points

Weighted Mean: $$\overline{x} = \frac{\sum (x_i/\sigma_i^2)}{\sum (1/\sigma_i^2)} \pm \frac{1}{\sqrt{\sum (1/\sigma_i^2)}}$$

$$\lim \sigma_i \to \sigma \qquad \overline{x} = \frac{1}{N} \sum x_i \pm \frac{\sigma}{\sqrt{N}}$$

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

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