Research Methods

Robert E Simpson

Singapore University of Technology & Design

robert_simpson@sutd.edu.sg

September 19, 2018

Week 2 Objectives

- To understand, quantify, and model the type of random variations that we often encounter in experimental studies
- To understand the difference between discrete and continuous random variables and give examples of each
- Calculate probabilities using the complement of known probabilities
- Integrate continuous probability distribution functions over different ranges to calculate the probability
- Use cumulative distribution functions to calculate probabilities for certain ranges
- Calculate probabilities using both continuous and discrete distributions
- Calculate the mean, variance and expectation value of a discrete distributions
- Calculate the expected mean and variance of Binary, Poisson, and Normal probability distributions
- Use the Binary, Poisson, and Normal probability distributions to calculate the probability of an occurence

In Week 2 Content

- Random variables
- Discrete and continuous distributions
- Permutations and combinations
- Common probability distributions functions
  - Binomial
  - Poisson
  - Gaussian

Statistical parameters

$\bar{x}$ The simple arithmetical average of a sample:

$$\bar{x} = \frac{1}{n} \sum_{0}^{n} (x_i)$$

μ The simple arithmetical average of the population:

$$\mu = \frac{1}{N} \sum_{0}^{N} (x_i)$$

$\langle x \rangle$ The *modelled* average of the population

The term *expectation* is used to denote that $\langle x \rangle$ is a **predicted value** and not a result of measuring the whole population.
We need a model of the distribution to obtain expected values.

[$\bar{x}$ can equal $\langle x \rangle$, when all values are equally distributed]

Random Variables

Random variable

Discrete random variable

Continuous random variable

Random Variables

Random variable a numerical value whose measured value can change for repeated experiments.

Discrete random variable a random variable with a finite set of real numbers over its range

Continuous random variable a random variable where the data can take any value in an interval

它全部可能取到的不相同的值是有限个或可列无限多个，也可以说概率1以一定的规律分布在各个可能值上。这种随机变量称为"离散型随机变量"

连续型随机变量是指如果随机变量X的所有可能取值不可以逐个列举出来，而是取数轴上某一区间内的任一点的随机变量

Concept Question 2.1

Identify the continuous random variables:

a) The time that an individual is logged onto the internet during a given week

b) The mean number of defective solder joints on a sample of circuit boards

c) The number of faulty transistors on a circuit board

d) The lifetime of a medical implant

e) Number of bits transmitted in error

f) The strength of a concrete sample

g) The number of flights arriving at Changi airport in any given hour

Probability

A random variable is used to describe the result of a measurement. Probability is used to quantify the chance that a measurement lies in some range of values.[1]

We usually express the probability of X having a value in some range using the following forms:

- $\mathrm{P}(X \in [\text{lower limit, upper limit}])$
- $\mathrm{P}(\text{lower limit} \leq X \leq \text{upper limit})$

[1] The probability is usually obtained from a model, or commonly estimated using **Relative Frequencies**. E.g. Repeat a measurement a large number of times, n, and calculate the proportion of measurements that fall in the range of interest.

Concept Problem 2.2

The following probabilities apply to the random variable X, which denotes the life in hours of fluorescent tubes:

- $P(X \leq 5000) = 0.1$
- $P(5000 < X \leq 6000) = 0.3$
- $P(X \geq 8000) = 0.4$

What is the probability that the tube's life-time is greater than 6000 hours?

Concept Problem 2.2

What is the probability that the tube's life-time is greater than 6000 hours?

$P(X > 6000) = 1.0 - P(X \leq 5000) - P(5000 < X \leq 6000)$

$P(X > 6000) = 1.0 - 0.1 - 0.3$

$P(X > 6000) = 0.6$

60% of the tubes have a lifetime greater than 6000 hours.

0.6 is the *complement* of $P(X \leq 6000) = 0.4$

40% of the have a life less than 6000 hours.

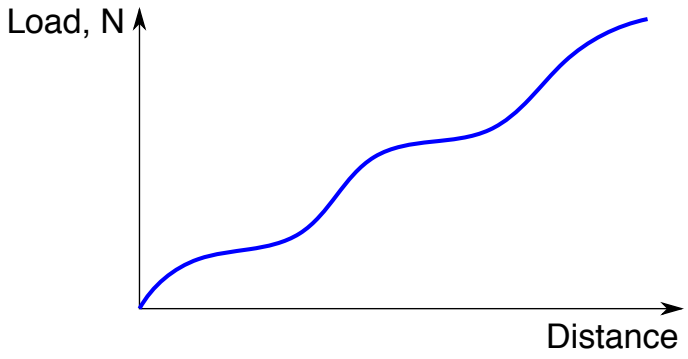
Distributions

Density Functions

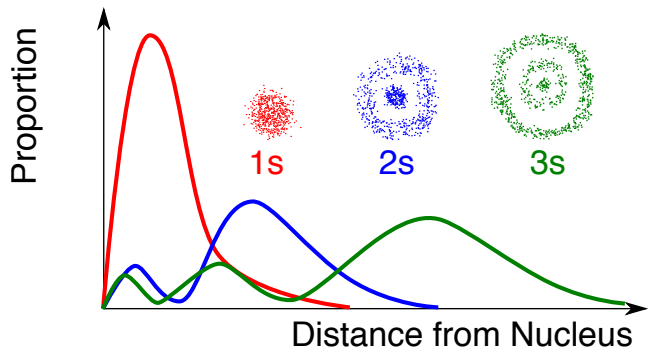
Density functions are used in engineering to describe how physical systems change vs a parameter of interest. E.g. Load as a function of length along a beam (a), or electron proportion as a function of radius from a nucleus (b), or the number of cars per a km along the ECP.

(a) At any point along the beam the load can be described as a density ($g\ cm^{-1}$). The total load between a and b is found by integrating the density between a and b.

(b) The proportion of electrons as a function of radius from the nucleus of a hydrogen atom. The probability of finding an electron between two different radii, a and b, is found by integrating the electron density function between a and b.



(a) Load vs distance



(b) e^- proportion vs radius

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Probability disitribution function (PDF)

Similarly, the PDF, $f(x)$, of a continuous random variable, X, describes the probability distribution of a continuous random variable.

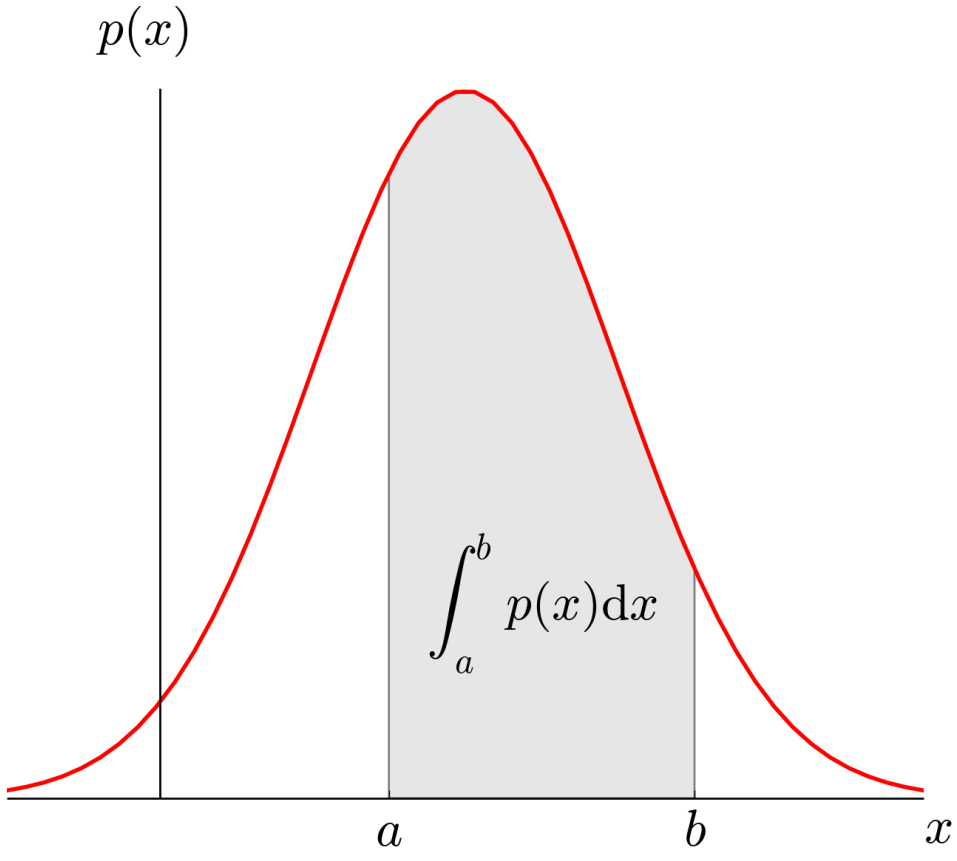
$$P(a < X < b) = \int_a^b f(x)dx \tag{1}$$

PDFs have the properties:

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

A histogram is an approximation to a PDF. The area of the bars represent the relative frequency (proportion) of the measurements in the interval

Probability distribution function (PDF)



Important: $f(x)$ is used to calculate an area that represents the probability that $a \leq X \leq b$.

Example 2.1

The continuous random variable X denotes the distance (μ m) from the start of a track on a magnetic disk until the first flaw. Historical data show that the distribution of X can be modelled by the pdf:

$$f(x) = \frac{1}{2000} e^{-\frac{x}{2000}}$$

a For what proportion of disks is the distance to the first flaw greater than 1000 μm?

b What proportion of the flaws are between 1000 and 2000 μm?

Example 2.1 Solution

(a)

$$P(X > 1000) = \int_{1000}^{\infty} f(x)dx$$

$$= \int_{1000}^{\infty} \frac{1}{2000} e^{-\frac{x}{2000}} dx$$

$$= e^{-\frac{1}{2}} = 0.607$$

(b)

$$P(1000 > X > 2000) = P(x > 1000) - P(x > 2000)$$

$$P(x > 2000) = \int_{2000}^{\infty} \frac{1}{2000} e^{-\frac{x}{2000}} dx$$

$$= e^{-1} = 0.368$$

$$P(1000 > X > 2000) = e^{-\frac{1}{2}} - e^{-1}$$

$$= 0.239$$

Common continuous distributions

Normal $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where the mean is μ and the variance is σ^2.
Whenever a random experiment is replicated to establish an average of value, and then this average is calculated many times for different samples, then the distribution of the average values tends to a normal distribution.

Lognormal $f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$, where the mean is μ and the variance is σ^2.
Variables sometimes follow an exponential relationship as $x = e^w$. If the exponent, w, is a random variable, then x is a random variable and follows a Lognormal distribution.

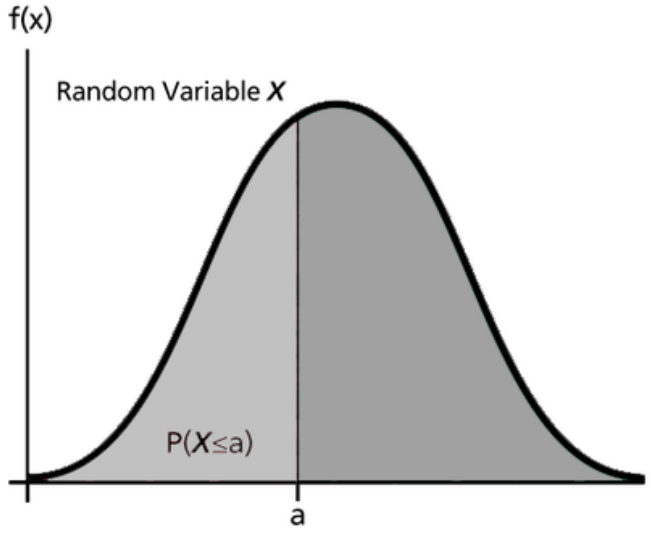
Weibull $f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^\beta}$, for $x > 0$. δ and β are scale and shape parameters respectively.
Used to model the time until failure of many different physical systems.

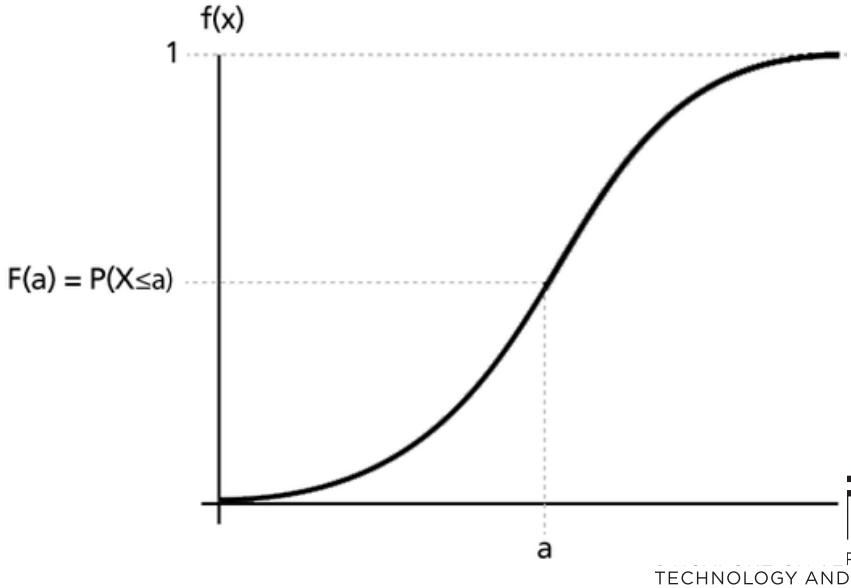
Cumulative distribution function (CDF)

The CDF, $F(x)$, of a continuous random variable, X, is used to determine the probability that X is less than or equal to a value x.

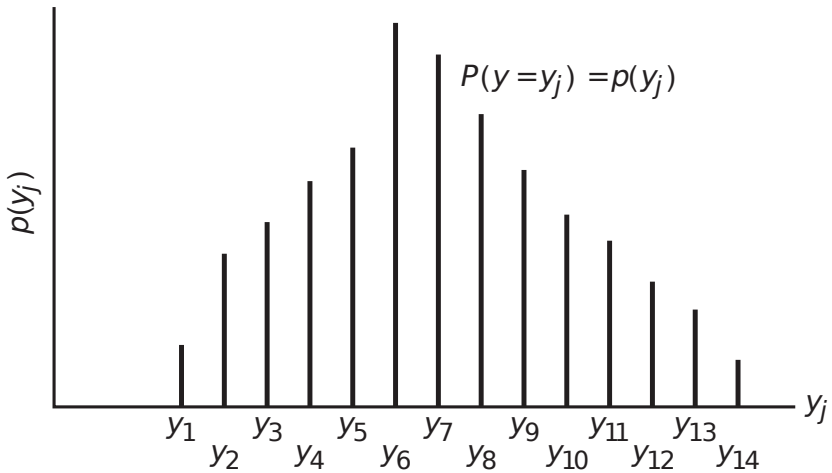
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du \tag{4}$$

for $-\infty < x < \infty$

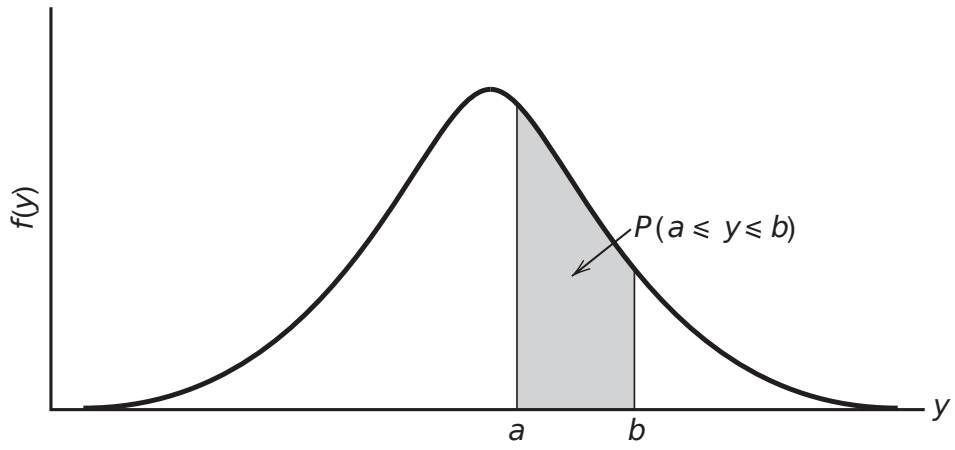




Discrete and Continuous Distributions



(*a*) A discrete distribution



(*b*) A continuous distribution

Discrete random variable Can only take a distinct value

Continuous random variable Can take any value

Concept Question 2.3

Is it possible to give a probability value to each individual value of a continuous random variable?

Concept Question 2.3

No, each value of a continuous variable may assume an infinite number of different values. Instead, probabilities are assigned to an interval of values.

Parameters of continuous distributions

For smoothly varying functions, we replace the sum over individual contributions by an integral over all values of x multiplied by the probability of x occurring.

$$\langle x \rangle = \int_{-\infty}^{\infty} x p(x) dx \tag{5}$$

Variance:

$$\sigma^2 = \frac{1}{N} \sum (x - \langle x \rangle)^2 = \bar{x}^2 - \langle x \rangle^2 \tag{6}$$

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 p(x) dx - \langle x \rangle^2 = \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 p(x) dx \tag{7}$$

Expectation Value:

$$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x) p(x) dx \tag{8}$$

Discrete Distributions

The Leicester City Story

1884	Leicester Fosse were founded
1919	Leicester Fosse were reformed as Leicester City
1929	Finished 2^{nd} in the top division
1949, 1961, 1963,1969	Reached the FA Cup final but lost
1964, 1997, 2000	Won the League Cup
2002	Leicester went into administration (bankrupt)
2008	Relegated to the 3^{rd} tier of English football
2009	Promoted back to the 2nd tier of English football
2014	Promoted to the Premier League
May 2015	Battled to escape relegation
Aug 2015	Leicester were 12-1 to be relegated and 5000-1 to win the league

What happened next?

5000/1: The impossible is possible

Leicester won their first ever Premiere League title in their 132 year history.

Why were the odds 5000/1?

- The analysts use the previous performance of all teams to assign a probability of each result
- Monte-Carlo Models are then used to predict the final league position for each club's final points tally
- The Monte Carlo simulations predict a distribution of final points tally. For a 100,000 trial MC simulation:
  - Leicester's mean points tally was 43 points with a standard deviation of 7 points
  - In the 100,000 trial MC simulation, Leicester only won the league once! The 5000/1 odds were generous.
- Generally, the number of goals/game follows a Poisson distribution, see here.

The models to predict scores did not account for Claudio Ranieri, Jamie Vardy, Ngolo Kante, Riyad Mahrez, Shinji Okazaki, and Danny Drinkwater's exceptional performances!

Case Problem 1: 63 Year Hurricane Stats

Number in one year	Occurences
2	1
3	8
4	11
5	9
6	8
7	10
8	6
9	5
10	1
11	2
12	1
15	1

(a) What is the mean number of hurricanes per a year?

(b) Using this distribution, estimate the probability of there being more than 10 hurricanes in any given year.

(a) Mean number of hurricanes:

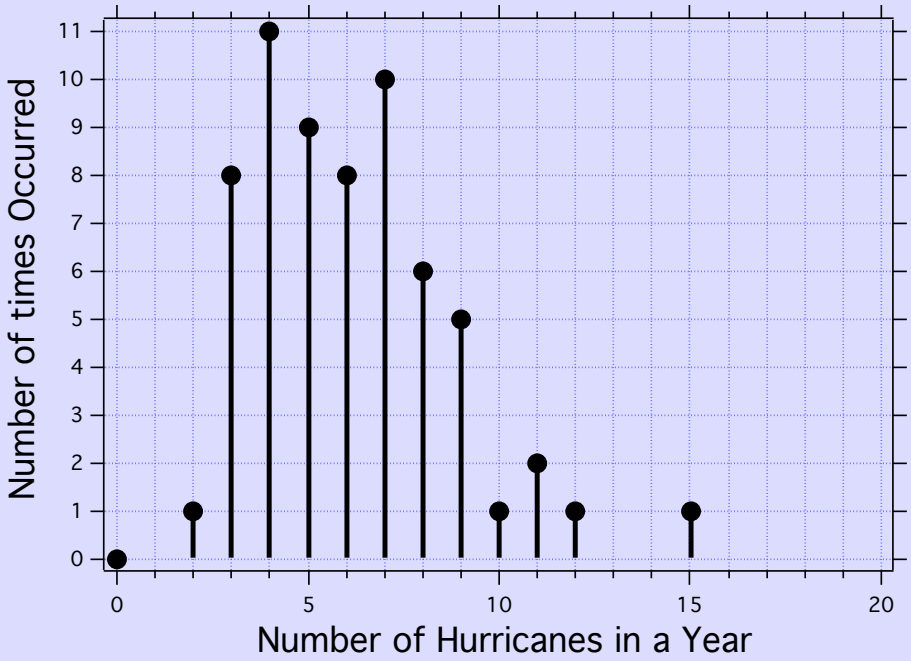
$$\bar{x} = \frac{(2 \times 1) + (3 \times 8) + ... + (15 \times 1)}{63} = 6.1$$

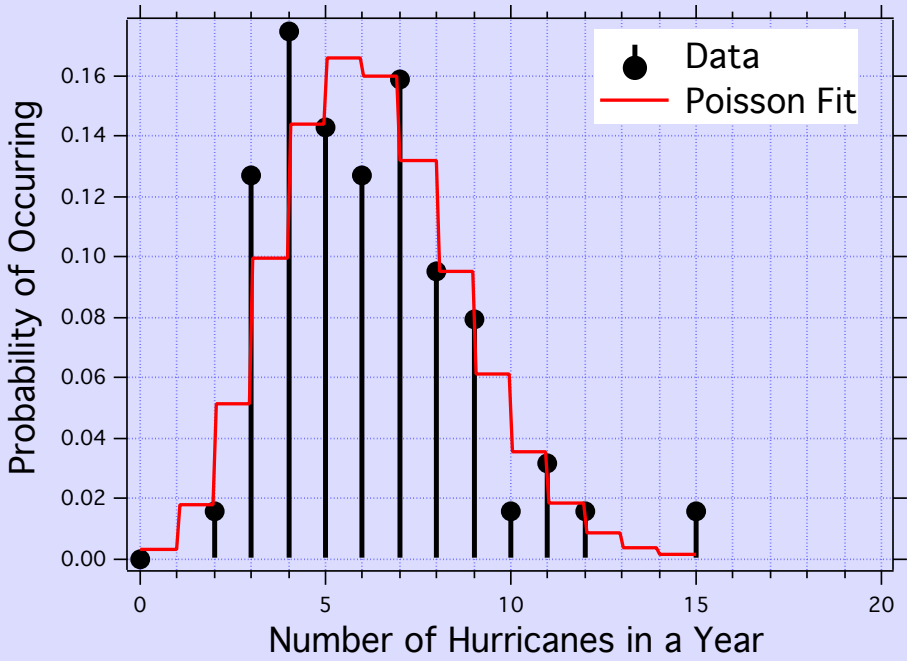
$$\bar{x} = (2 \times \frac{1}{63}) + (3 \times \frac{8}{63}) + ... + (15 \times \frac{1}{63}) = 6.1$$

$$\bar{x} = \sum_{j}^{n} x_j P(x_j) \quad (9)$$

(b) Estimate of more than 10 hurricanes: This only happened in 4 of the 63 years. Therefore, $P(N > 10) = 4/63 = 0.063$. A 6% chance

Frequency and Probability of Hurricanes





SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Mean, variance and expectation value of discrete distributions

Sample Mean:

$$\boxed{\bar{x} = \sum_{j}^{n} x_j P(x_j)} \quad \text{and therefore} \quad \bar{x^2} = \sum_{j}^{n} x_j^2 P(x_j) \tag{10}$$

Variance[2]**:**

$$\boxed{\sigma^2 = \bar{x^2} - \mu^2 = \sum_{j=1}^{n} [x_j^2 P(x_j)] - \mu^2} \tag{11}$$

Expectation Value: The expectation value of any function, *f(x)*, is given by:

$$\boxed{\langle f(x) \rangle = \lim_{N \to \infty} \sum_{j}^{N} [f(x_j) P(x_j)]} \tag{12}$$

[2] $P(x_j)$ is also the probability that x_j and therefore $(x_j - \mu)$ and x_j^2 will occur, i.e. $\sigma^2 = \lim_{n\to\infty} \sum_{j=1}^{n} [(x_j - \mu)^2 P(x_j)]$

Binomial Distributions

Conditions of a binomial experiment

1. There are n identical trials
2. Each trial has only two possible outcomes (success or failure)
3. The probability outcomes is constant for each trial
4. The trials are independent

p= probability of success
q= probability of failure =1-p

Binomial Experiments

Often used in experiments that measure proportions

- Poling 1000 people if they have ever been to the King Power Stadium.
- Rolling two dice to see if you get a double
- Firing a beam of $\alpha-$particles at foils of gold leaf, and counting the number of atoms that are defected by an angle $> 90^{\circ}$

What is the chance of getting 3 heads from 5 coin tosses?

What is the chance of getting 3 heads from 5 coin tosses?

Probability of Getting 5 heads:

Permutation	Coin 1	Coin 2	Coin 3	Coin 4	Coin 5	P
1	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$(\frac{1}{2})^5 = \frac{1}{32}$

Probability of Getting 4 heads:

Permutation	Coin 1	Coin 2	Coin 3	Coin 4	Coin 5	P
1	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	$\frac{1}{32}$
2	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
3	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
4	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
5	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$

We do not care about the order of the coins, therefore
$P(H = 4) = 5 \times \frac{1}{32} = \frac{5}{32}$

What is the chance of getting 3 heads from 5 coin tosses?

Probability of Getting 3 heads:

Permutation	Coin 1	Coin 2	Coin 3	Coin 4	Coin 5	P
1	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	T $\frac{1}{2}$	$\frac{1}{32}$
2	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	$\frac{1}{32}$
3	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
4	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	$\frac{1}{32}$
5	H $\frac{1}{2}$	T $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
6	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
7	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	$\frac{1}{32}$
8	T $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
9	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$
10	T $\frac{1}{2}$	T $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	H $\frac{1}{2}$	$\frac{1}{32}$

We do not care about the order of the coins, therefore
$P(H = 3) = 10 \times \frac{1}{32} = \frac{10}{32}$

排列 Permutations

Consider a set with 3 elements $S = (a, b, c)$. The number of different ways we can arrange these elements (number of permutations) is $n!$. i.e. for S, $n = 3$, and we have 6 permutations: abc, acb, bac, bca, cab, and cba. The number of permutations of a set of n elements is:

$$\boxed{n! = n \times (n-1) \times (n-2) \times (n-3) \times \cdots \times 2 \times 1}$$

If we are only interested in the arrangement of x of n elements, it follows that:

$$\boxed{P_m(n, x) = n \times (n-1) \times (n-2) \times \cdots \times (n-x+1) = \frac{n!}{(n-x)!}}$$

Combinations

Often we are only interested different subsets of x from a set of n elements. Each subset has a different value, but the elements in the subset can be arranged in different orders. For example for the set $S = (a, b, c, d)$, a subset might be abc, acb, bac, bca, cab, and cba, all of which have the same value.

The number of combinations, subsets of x elements that can be selected from a set of n elements, is:

$$C(n, x) = \frac{P_m(n, x)}{x!} = \frac{n!}{(n-x)!x!} = \binom{n}{x}$$

Example: Permutations and combinations of coins

Number of different ways that 5 coins can be arranged:

$$P_m(n, x) = n.(n-1)(n-2)...(n-x+1)$$
$$P_m(5, 5) = 5.4.3.2.1 = 120$$

Number of different ways to select 3 coins from 5:

$$P_m(5, 3) = 5.4.3 = 60$$

However, we have degeneracy (several combinations look the same– the position doesn't matter).
We have 3 coins, how many ways can they be arranged?
1^{st} coin can exist in 3 spaces, 2^{nd} coin can exist in 2 spaces, and 3^{rd} coin, can only exist in 1 place.
Therefore 3! ways of arranging 3 coins, or generally $x!$.

The number groups with different numbers of heads:

$$\frac{P_m}{x!} = \frac{n!}{(n-x)!x!} \tag{13}$$

We then can multiply the number of different groups by $\frac{1}{32}$ to find the likelihood that group occurring.

Permutations and combinations

Number of ways to arrange n coins: $n!$

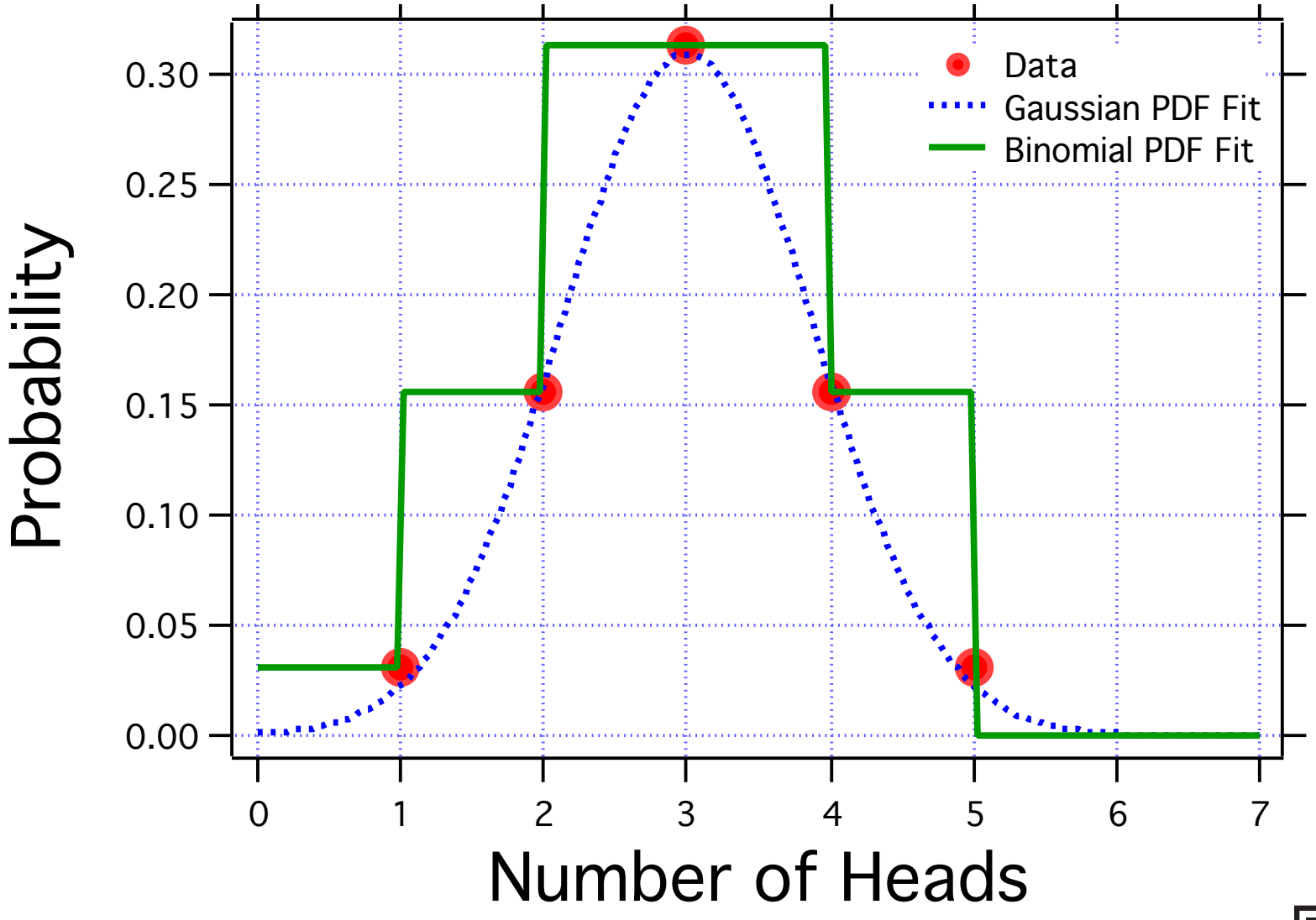
Number of ways to arrange x coins from a total of n coins:

$$P_m(n,x) = \frac{n!}{(n-x)!} \tag{14}$$

Number of distinct combinations:

$$\boxed{C(n,x) = \frac{P_m(n,x)}{x!} = \frac{n!}{(n-x)!x!} = \binom{n}{x}} \tag{15}$$

Probability distribution for 5 coins



Probability of combinations

The probability that we get x heads and (n-x) tails is the probability that combination will occur (P=$(\frac{1}{2})^n$) multiplied by the number of different ways to achieve that particular combination of x heads and n-x tails.

More generally, if the chances are probability p for one event and q=1-p for the other event,

$$P(x, n, p) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad (16)$$

Case Problem 2.2

What is the probability of at least 2 people in the class sharing the same birthday?

(a) How many different pairs can we form (allow multiple pairs to have the same birthday)?

(b) Why can't we use equation 16 for this problem?

(c) What is the probability that two randomly selected people do not have the same birthday?[3]

(d) What is the probability that any of possible pairs contain two people with the same birthday?

[3] You may assume that the daily birth rate is uniformly distributed throughout the year.

Case problem 2.2- Happy birthday

(a) 30 people in the class (n=30). Therefore, the number of ways that people can combine to make a pair (x=2) is:

$$C(n,x) = \frac{n!}{(n-x)!x!}$$

$$C(30) = \frac{30!}{(30-2)!2!} = 435$$

(b) The binomial approximation assumes independence. I.e. that one birthday pair does not affect another birthday pair. But this is not true, if Person 1 and Person 3 match, and Person 3 and 5 match, we know that 1 and 5 match also. The outcome of 1 and 5 depends on their results with 3, which means the results aren't an independent 1/365 chance.

(c) To make the problem easier, consider the chance of 2 people having **different** birthdays: $P_{diff} = \frac{364}{365}$.
(d) There are 435 different pairs of people, therefore the chance of two people **not** sharing a birthday is raised to the power of 435.

$$P_{diff} = (\frac{364}{365})^{435} = 0.3$$

So the chance of two people in the class sharing the same birthday is 70%.

Mean and Variance of a Binomial Distribution

Mean:

$$\langle x \rangle = \sum_{x=0}^{x=\infty} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$\boxed{\langle x \rangle = np}$$

Variance:

$$\boxed{\sigma^2 = np(1-p)}$$

(The derivations may be included as H/W)

Case Problem 2.3

A particle physicist measures the angular distribution of K-mesons scattered from a liquid target. There should be equal numbers of particles scattered forwards and backwards. After 1000 measurements, 472 are scattered forwards, 528 are scattered backwards.

(a) What uncertainty should she give assuming the theoretical mean is known?

(b) What uncertainty should she give when the theoretical mean is unknown?

Case Problem 2.3-Solution

What uncertainty should she give assuming the theoretical mean is known?

$$\sigma^2 = np(1-p) \quad \text{For binomial distribution}$$
$$\sigma^2 = 1000 \times 0.5(1-0.5)$$
$$\sigma = 15.81$$

What uncertainty should she give when the theoretical mean is unknown?

$$S^2 = np(1-p)$$
$$S^2 = 1000 \times 0.472(0.528)$$
$$S = 15.78$$

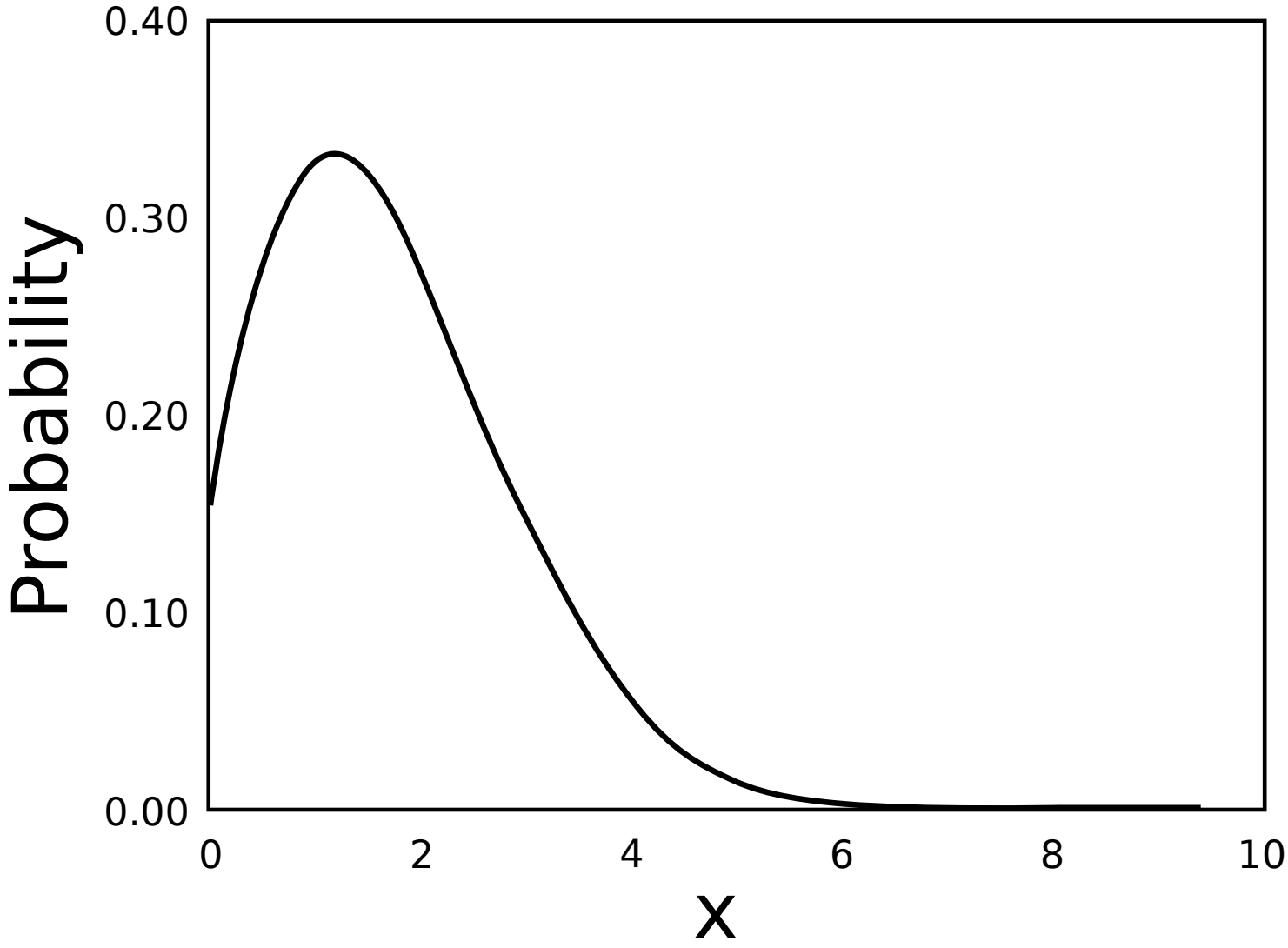
NOTE: when P is close to 50%, σ is insensitive to uncertainty in determining P.

Poisson Distribution

- Used to model the probability of occurrences in counting experiments
- An approximation to the binomial distribution when $\mu \ll n$ due to $p \ll 1$
  (For large n, approximate Binomial to Poisson to get rid of factorials.)
- Can be applied even when we do not know n and p (instead we use the mean count rate)
- For a random variable, x, Poisson distributions have the nice property that $E(X) = Var(X) = \mu$

$$\boxed{p(x, \mu) \equiv \frac{\mu^x}{x!} e^{-\mu}} \tag{17}$$

Poisson Distribution



Proof that $\langle x \rangle = \mu$ in Poisson distributon

The parameter, μ, in the poisson distribution function is the expectation value of a random variable X

$$\langle X \rangle = \sum_{i=0} x_i p(x_i) \tag{18}$$

$$= \sum_{i=0} x_i \frac{e^{-\mu} \mu^{x_i}}{x_i!} \tag{19}$$

$$= \sum_{i=0} x \frac{e^{-\mu} \mu \mu^{x_i - 1}}{x_i (x_i - 1)!} = \mu e^{-\mu} \sum_{i=0} \frac{\mu^{x_i - 1}}{(x_i - 1)!} \tag{20}$$

$$\tag{21}$$

let $k = x_i - 1$, so we can make the Taylor series: $e^k = \sum_{k=0} \frac{\mu^k}{k!}$, then:

$$\langle X \rangle = e^{-\mu} \mu \sum_{k=0} \frac{\mu^k}{k!} = e^{-\mu} \mu e^{\mu} = \mu \tag{22}$$

$$\tag{23}$$

Mean and Variance of a Poisson Distribution

The mean is:

$$\langle X \rangle = \mu \tag{24}$$

The variance is:

$$\sigma^2 = \mu \tag{25}$$

This result is extremely useful when planning measurement times for an experiment, why?

Mean and Variance of Poisson a Distribution

The mean is:

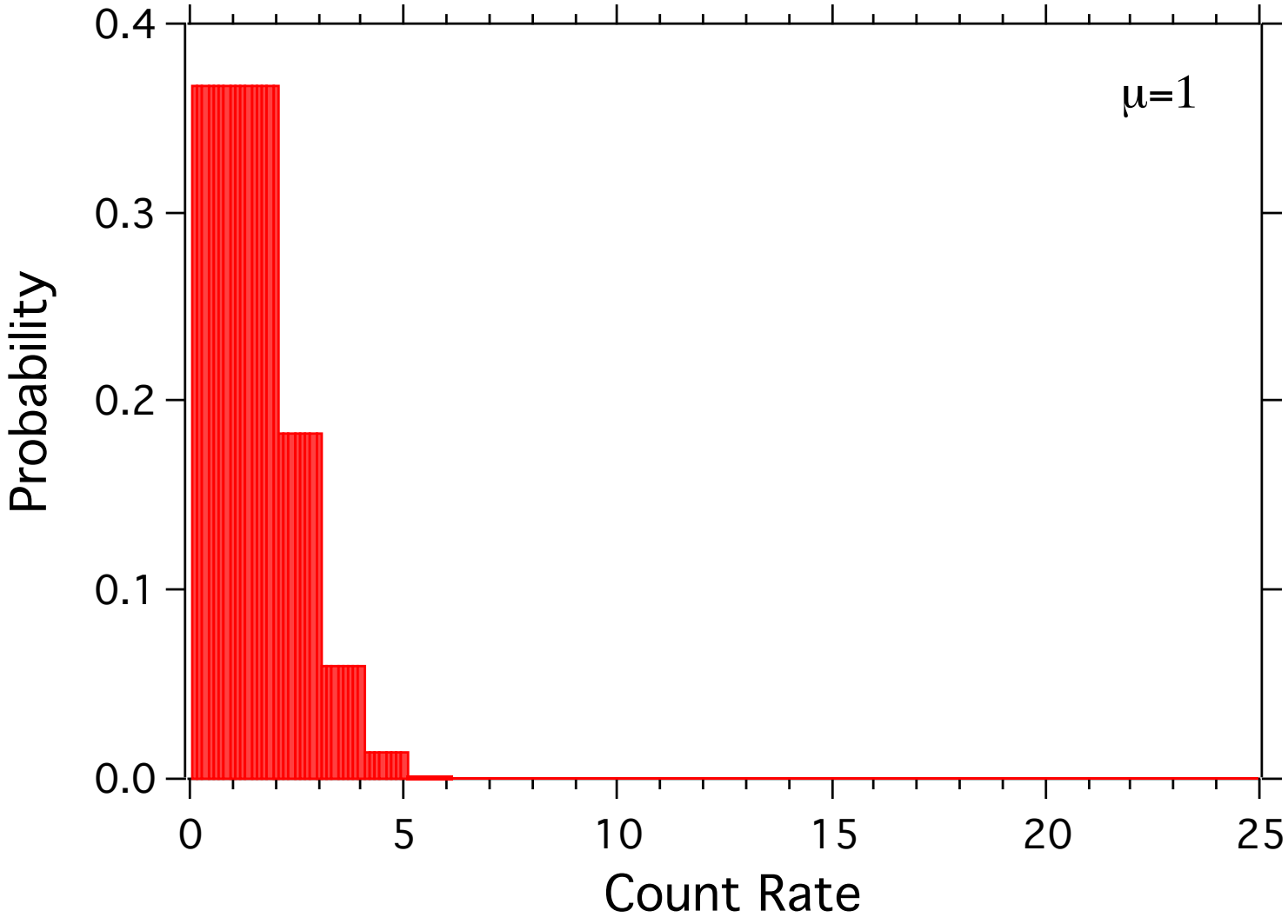
$$\langle x \rangle = \mu$$

The variance is:

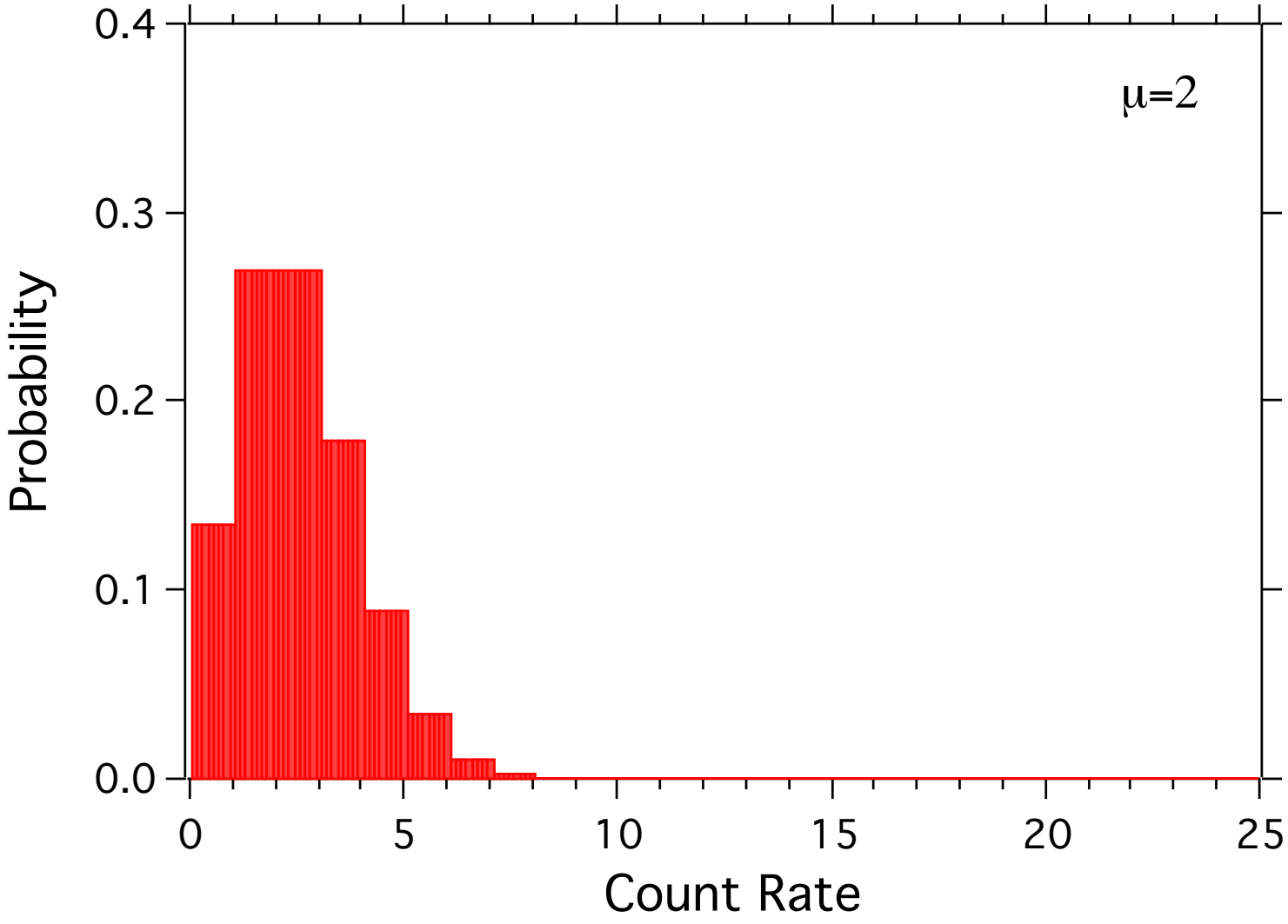
$$\sigma^2 = \mu$$

A very useful result when designing counting experiments. We can reduce the measurement uncertainty by collecting data for longer and therefore getting a larger number of counts. E.g. Assume we have a mean count rate of 1 count per second. If we conduct a 1 second experiment, we would expect to measure 1 count. Therefore the standard deviation is $\sqrt{1} = 1$. i.e. the spread of data is as wide as the mean the data. Now suppose, we increase the measurement time by a factor of four to 4 sec, then on average we measure 4 counts ($\mu = 4$ with a standard deviation $\sqrt{4} = 2$. Therefore the spread in the data has been halved but the experiment took 4 times longer! Generally for Poisson statistics the Signal-to-Noise ratio scales as $\sqrt{(time)}$.

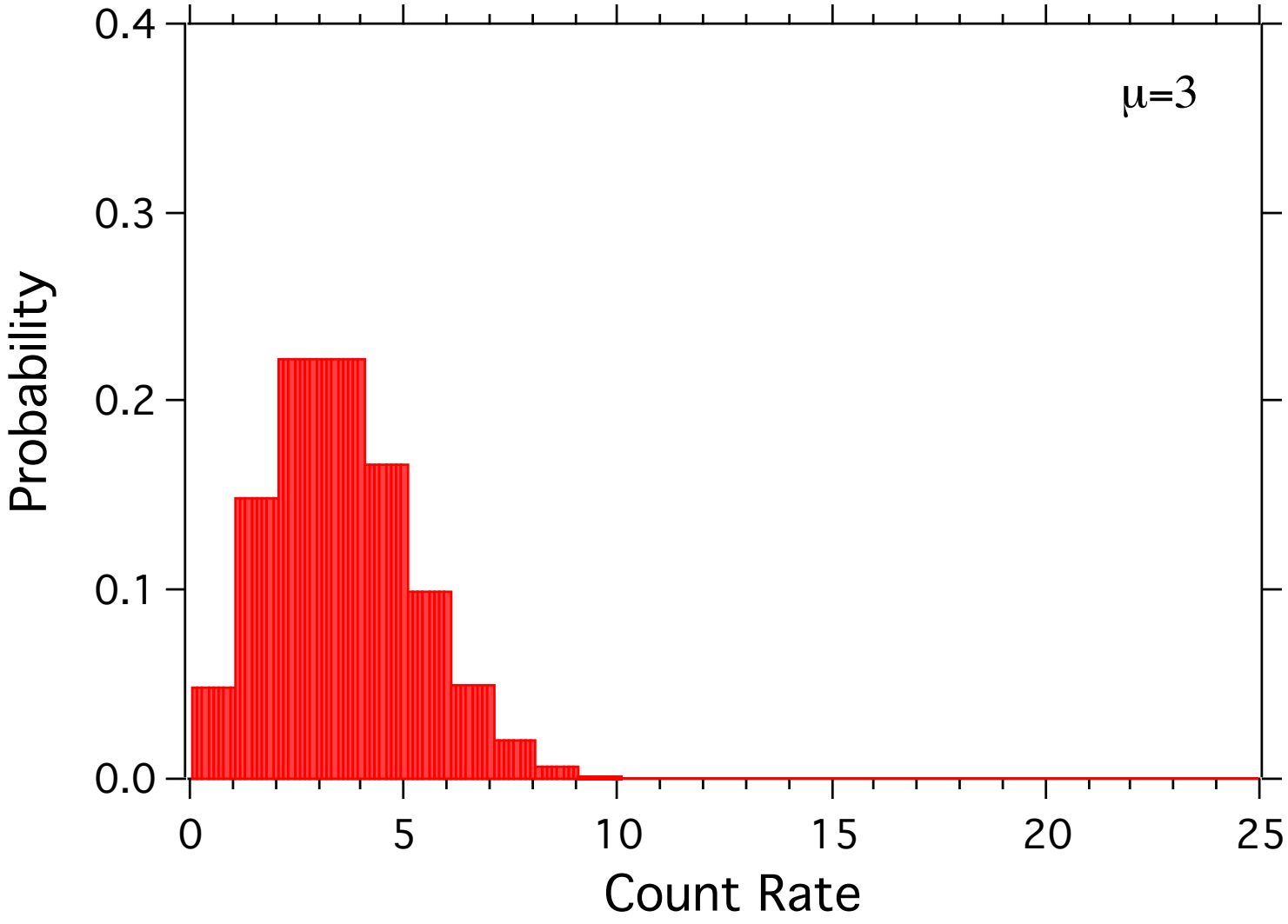
Poisson Distribution Shape Depends on μ



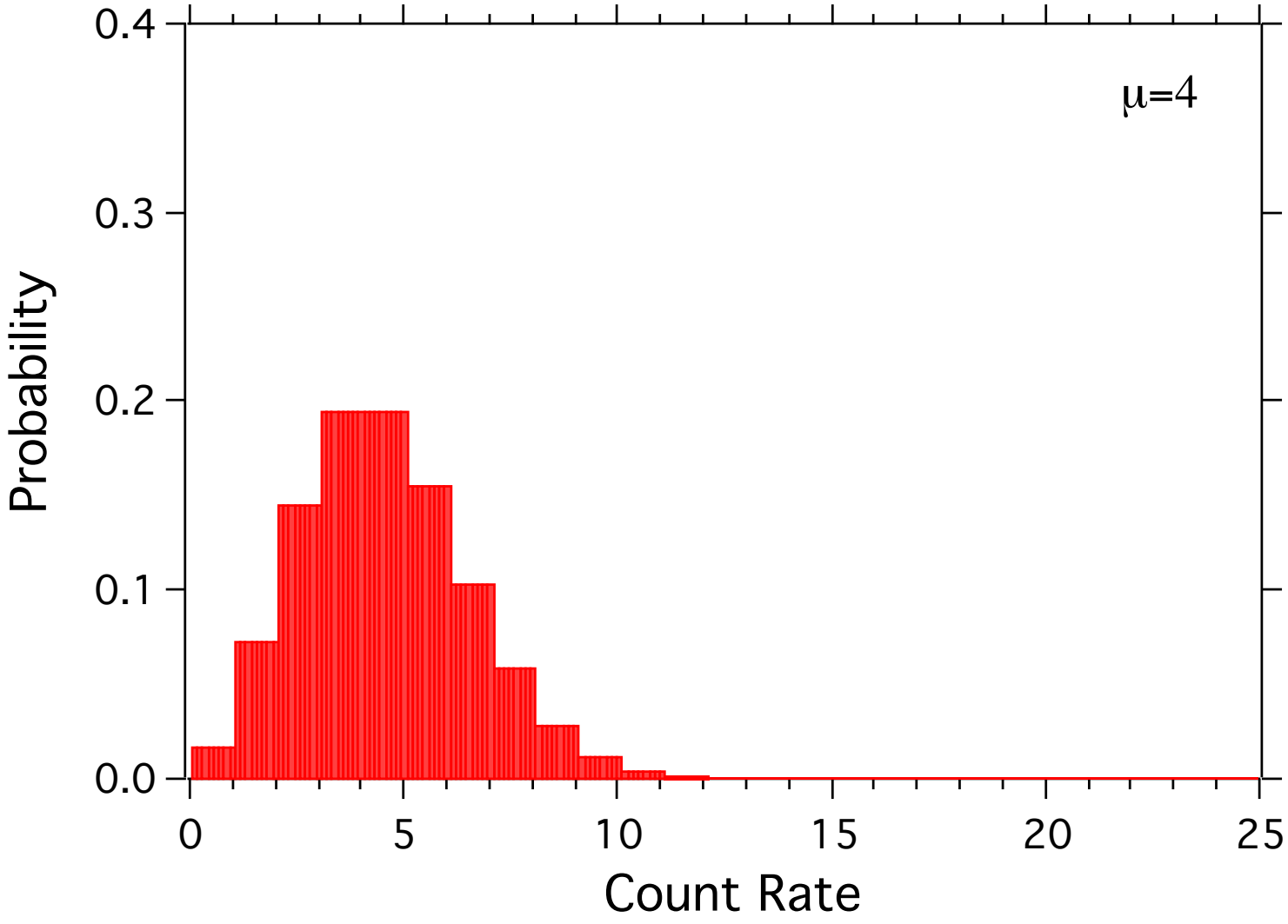
Poisson Distribution Shape Depends on μ



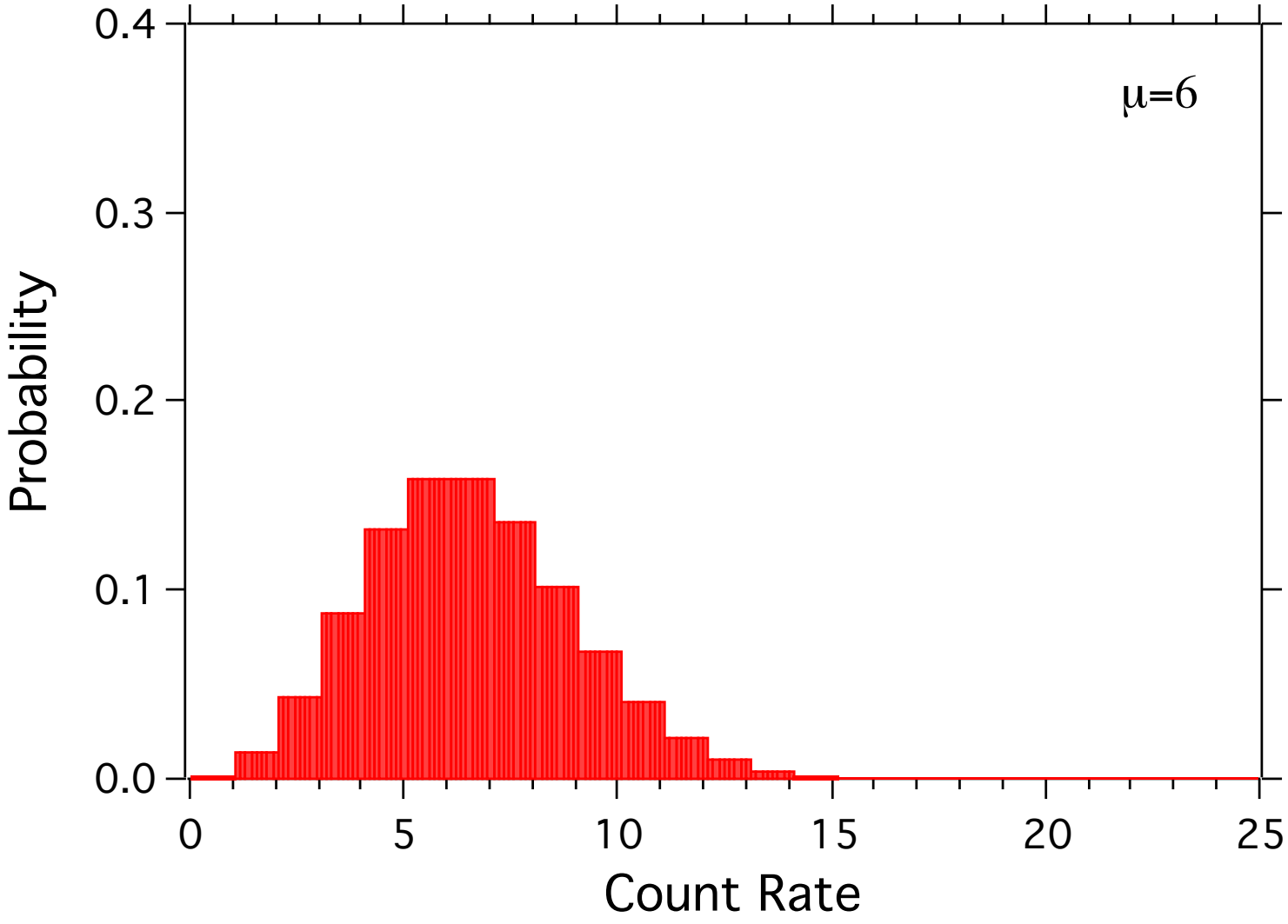
Poisson Distribution Shape Depends on μ



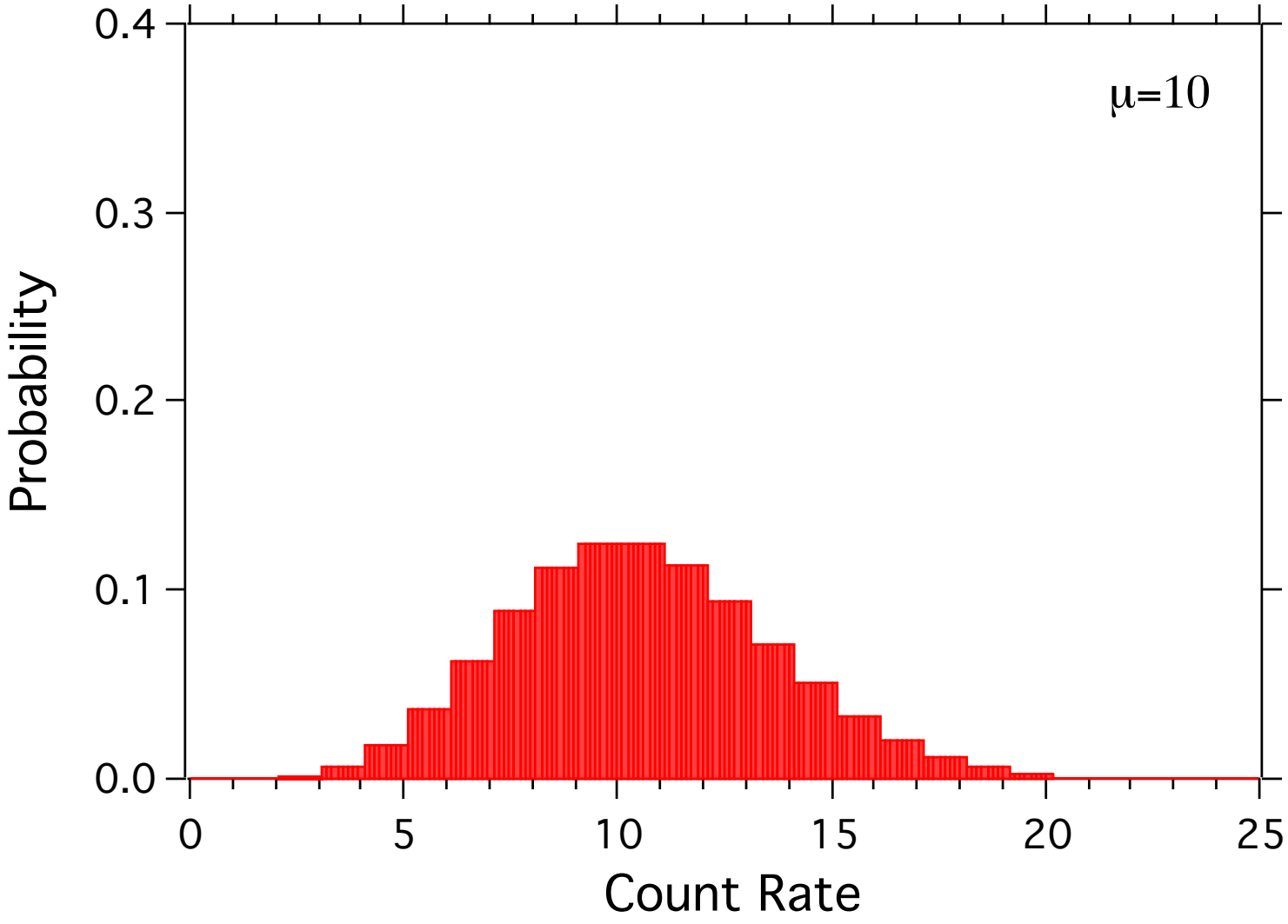
Poisson Distribution Shape Depends on μ



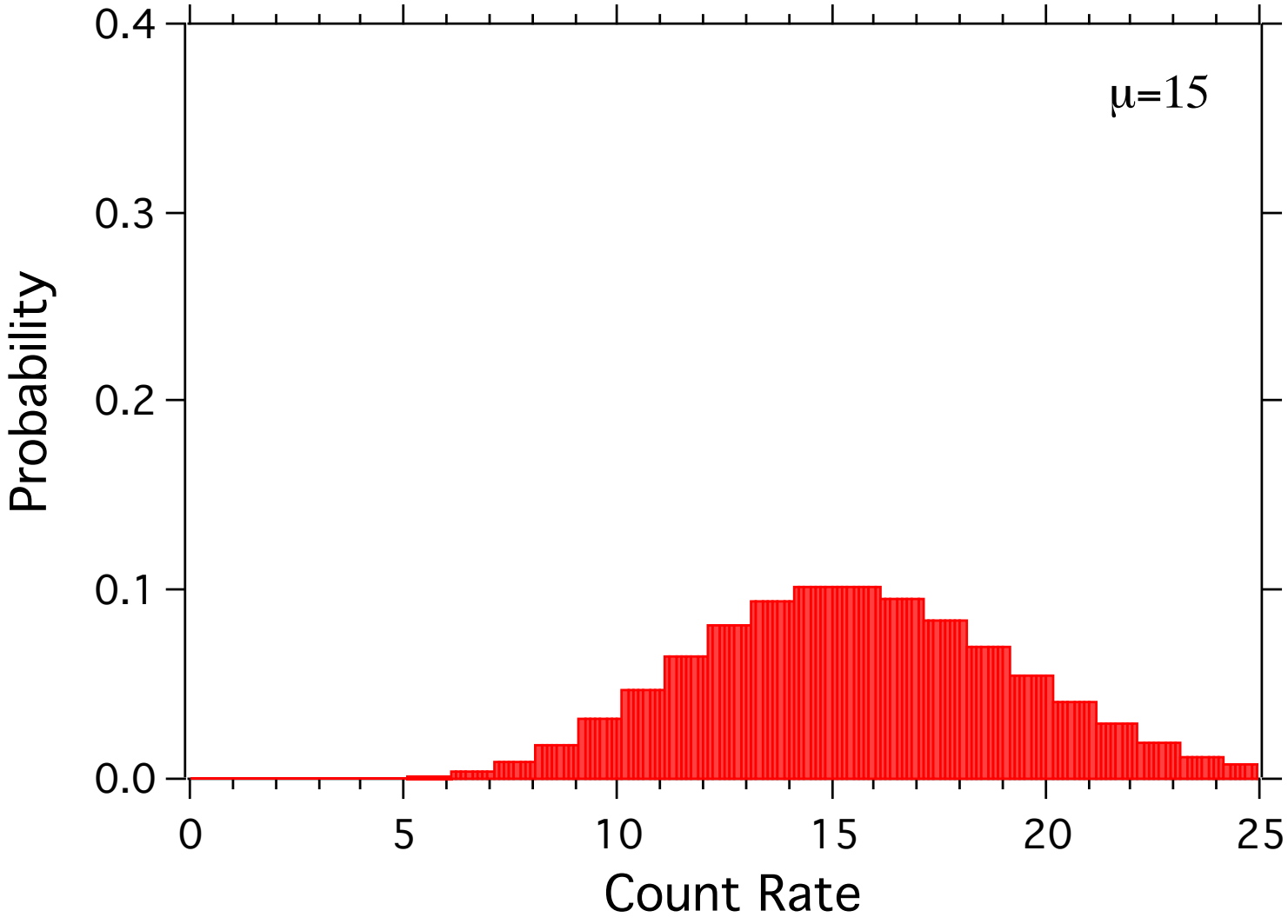
Poisson Distribution Shape Depends on μ



Poisson Distribution Shape Depends on μ



Poisson Distribution Shape Depends on μ



Case Problem 2.4

The number of pinholes in sheets of plastic are of concern to a manufacturer. If the number of pinholes is too large, the plastic is unusable. The mean number of pinholes per square metre is equal to 1.5. The 1 m^2 sheets are unusable if the number of pinholes exceeds 4. If 50 sheets are manufactured, how many sheets will be unusable?

Case Problem 2.4 Solution

$$\boxed{p(x,\mu) = \frac{\mu^x}{x!}e^{-\mu}}$$

$$p(0,1.5) = \frac{1.5^0}{0!}e^{-1.5} = 0.223$$

$$p(1,1.5) = \frac{1.5^1}{1!}e^{-1.5} = 0.335$$

$$p(2,1.5) = \frac{1.5^2}{2!}e^{-1.5} = 0.251$$

$$p(3,1.5) = \frac{1.5^3}{3!}e^{-1.5} = 0.126$$

$$p(4,1.5) = \frac{1.5^4}{4!}e^{-1.5} = 0.047$$

$$p(x \leq 4) = p(0,1.5) + p(1,1.5) + p(2,1.5) + p(3,1.5) + p(4,1.5) = 0.981$$

$$p(x > 4) = 1 - p(x \leq 4) = 0.0185$$

Therefore 1.9% of manufactured sheets of unusable, which is ~ 1 sheet in 50.

Normal Distribution

The normal distribution is often observed in nature because it naturally forms when one analyses distribution of averages of any distribution. I.e. a sample distribution.

A binomial distribution approximates to a normal distribution when:

- $n \to \infty$
- $np \gg 1$

It is also the limiting case of the Poisson distribution when n becomes large.

The Normal/Gaussian PDF is:

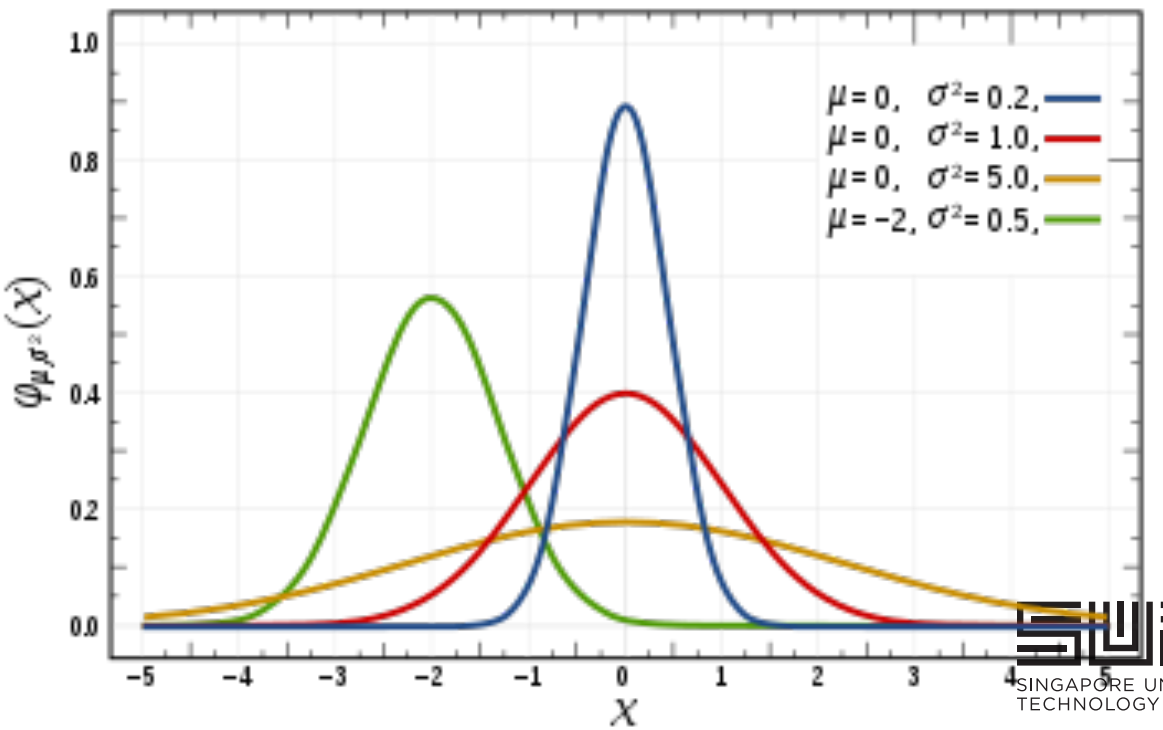
$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{26}$$

x= random variable

μ = mean

σ^2 = variance

σ = standard deviation



Standard Gaussian Distribution

Define a dimensionless variable: $z = \frac{x-\mu}{\sigma}$

$$\boxed{p_G(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)} \tag{27}$$

Useful when looking up values in tables or designing an efficient computer code to find $p_G(z)$. Simply, calculate z and then search for $p_G(z)$ then scale the value by $\frac{1}{\sigma}$.

Summary

Discrete Distributions:
Sample Mean: $\bar{x} = \sum_j^n x_j P(x_j)$
Variance: $\sigma^2 = \sum_{j=1}^n [x_j^2 P(x_j)] - \mu^2$
Expectation Value: $\langle f(x) \rangle = \lim_{N\to\infty} \sum_j^N [f(x_j)P(x_j)]$

Continuos Distributions:
Mean: $\mu = \int_{-\infty}^{\infty} xp(x)dx$
Variance: $\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx = \int_{-\infty}^{\infty} x^2 p(x)dx - \mu^2$
Expectation Value: $\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)p(x)dx$

Permutations and Combinations
Permutations: $P_m(n,x) = \frac{n!}{(n-x)!}$
Combinations: $C(n,x) = \frac{P_m(n,x)}{x!} = \frac{n!}{(n-x)!x!}$

Binomial distribution
mean: $\langle x \rangle = np$
variance: $\sigma^2 = np(1-p)$
Poisson distribution
PDF: $P_p(x,\mu) \equiv \frac{\mu^x}{x!} \exp^{-\mu}$
Mean: $\langle x \rangle = \mu$
Variance: $\sigma^2 = \mu$

Gaussian distribution
$p_G = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$

Recap Presenter Schedule

DATE	ID	STUDENT		
12/9/18	1003281	Zhang Rui	PHD	EPD
18/9/18	1003967	Rajendran Meena	PHD	EPD
19/9/18	1000246	Chau Zhong Hoo	PHD	EPD
25/9/18	1003968	Luo Yin-Jyun	PHD	ISTD
26/9/18	1000378	Liu Junhua	PHD	ISTD
2/10/18	1003971	Manivannan Ajaykumar	PHD	EPD
3/10/18	1000544	Kwa Hian Lee	PHD	EPD
9/10/18	1004028	Liu Bowen	PHD	ISTD
10/10/18	1000583	Lee Cheng Pau	PHD	EPD
16/10/18	1004029	Xu Xiansong	PHD	SCI
17/10/18	1000949	Khairuldanial Bin Ismail	PHD	EPD
30/10/18	1004031	Chen Shaohua	PHD	EPD
31/10/18	1001033	Tan Yeh Wen	MASTER	EPD
6/11/18	1004032	YEO SUE-MAE	PHD	ISTD
7/11/18	1001175	Koh Zann	PHD	EPD
13/11/18	1004036	Suhaila Binte Zainal Shah	PHD	HASS
14/11/18	1003270	Zhang Wang	PHD	SCI
20/11/18	1004037	Oh Peng Ho (Hu Binghe)	PHD	HASS
21/11/18	1003272	Kamya Nagarajan	PHD	EPD
27/11/18	1003278	Jia Yin	PHD	EPD
28/11/18	1003273	Ng Hsien Han	PHD	SCI
4/12/18	1003284	Shermaine Yvonne Tan	PHD	EPD
5/12/18	1003275	Chadurvedi Venkatesan	PHD	EPD
SPECIAL	1000974	Yeow Lih Wei	MASTER	ESD
SPECIAL	1003957	Omkar	PHD	EPD
SPECIAL	1003958	Feng Xiaolong	PHD	SCI
SPECIAL	1003959	Duraisamy Sasirekha	PHD	EPD
SPECIAL	1003960	Li Guangtong	PHD	EPD
SPECIAL	1003961	Rayudu Nithin Manohar Chowdary	PHD	EPD
SPECIAL	1003965	Shawndy Michael Lee Jin Lun	PHD	EPD
SPECIAL	1003966	Ng Shiwei	PHD	EPD
SPECIAL	1004030	Surovi Nowrin Akter	PHD	EPD