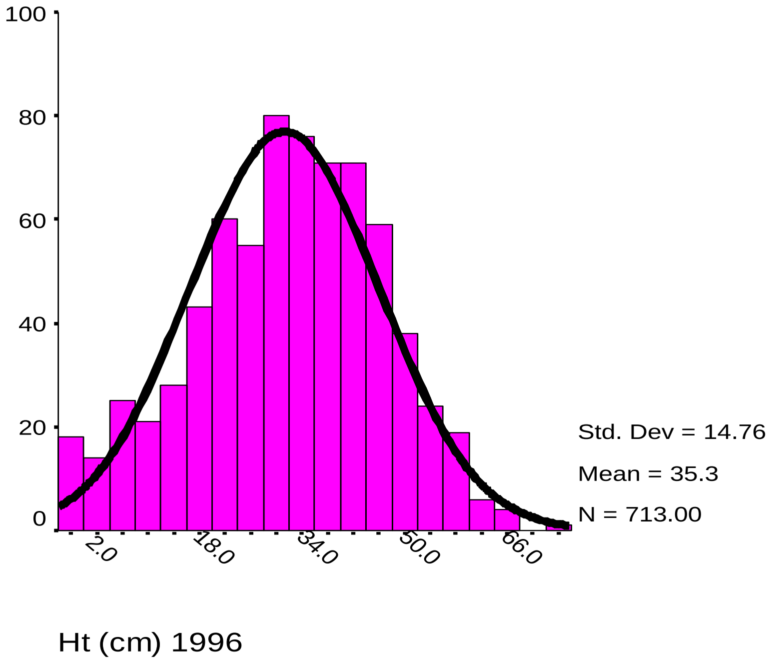
BT 623 Research Methodology

Probability Distribution

Prof. Utpal Bora
Department of Biosciences and Bioengineering Indian Institute of Technology Guwahati Kamrup, Assam- 781039, India
Email: ubora@ iitg.ac.in

Probability distributions

- We use probability distributions because they work –they fit lots of data in real world

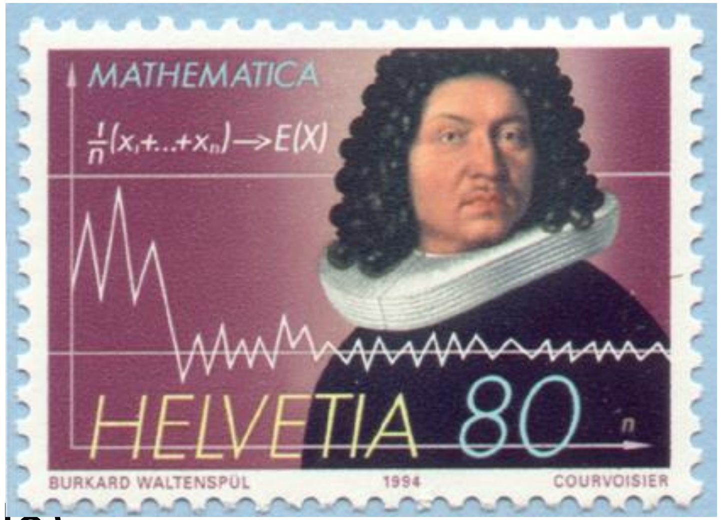


Height (cm) of *Hypericum cumulicola* at Archbold Biological Station

The Binomial Distribution

Bernoulli Random Variables

- Imagine a simple trial with only two possible outcomes
  - Success (*S*)
  - Failure (*F*)



Jacob Bernoulli (1654-1705)

- Examples
  - Toss of a coin (heads or tails)
  - Sex of a newborn (male or female)
  - Survival of an organism in a region (live or die)

The Binomial Distribution
Overview

- Suppose that the probability of success is p
- What is the probability of failure?
  - $q = 1 – p$
- Examples
  - Toss of a coin (S = head): $p = 0.5 \Rightarrow q = 0.5$
  - Roll of a die (S = 1): $p = 0.1667 \Rightarrow q = 0.8333$
  - Fertility of a chicken egg (S = fertile): $p = 0.8 \Rightarrow q = 0.2$

The Binomial Distribution
Overview

- Imagine that a trial is repeated *n* times

- Examples
  - A coin is tossed 5 times
  - A die is rolled 25 times
  - 50 chicken eggs are examined

- Assume *p* remains constant from trial to trial and that the trials are statistically independent of each other

The Binomial Distribution
Overview

- Formula for Binomial Distribution

$$P(x) = {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x}$$

$$= \frac{n!}{x!(n-x)!} p^{x} \cdot q^{n-x}$$

where ${}^{n}C_{x}$ is the number of ways to obtain x successes

Question 1: Find the binomial distribution of getting a six in three tosses of an unbiased dice.

Question 1: Find the binomial distribution of getting a six in three tosses of an unbiased dice.

Let X be the random variable of getting six. Then X can be 0, 1, 2, 3.

Here, $n = 3$

p = Probability of getting a six in a toss = $\frac{1}{6}$

q = Probability of not getting a six in a toss = $1 - \frac{1}{6} = \frac{5}{6}$

$P(X = 0) = {}^nC_r\, p^r\, q^{(n-r)} = {}^3C_0\, (\frac{1}{6})^0\, (\frac{5}{6})^{3-0} = 1 \times 1 \times 125/216 = 125/216$

$P(X = 1) = {}^nC_r\, p^r\, q^{(n-r)} = {}^3C_1\, (\frac{1}{6})^1\, (\frac{5}{6})^{3-1} = 3 \times \frac{1}{6} \times 25/36 = 25/72$

$P(X = 2) = {}^nC_r\, p^r\, q^{(n-r)} = {}^3C_2\, (\frac{1}{6})^2\, (\frac{5}{6})^{3-2} = 3 \times 1/36 \times \frac{5}{6} = 5/72$

$P(X = 3) = {}^nC_r\, p^r\, q^{(n-r)} = {}^3C_3\, (\frac{1}{6})^3\, (\frac{5}{6})^{3-3} = 1 \times 1/216 \times 1 = 1/216$

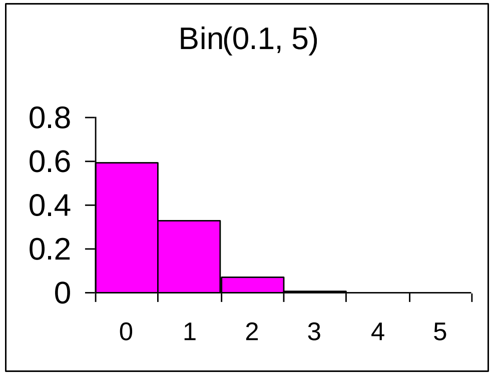
Question 2: When an unbiased coin is tossed eight times, what is the probability of obtaining:

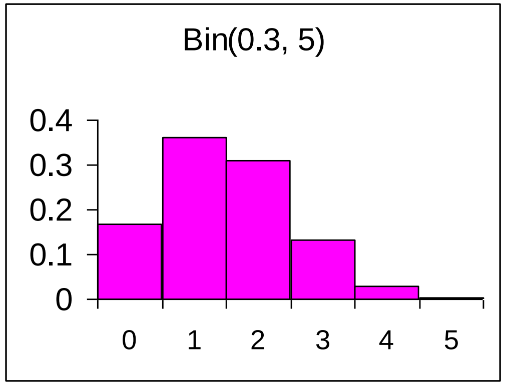
(a) Less than 4 heads

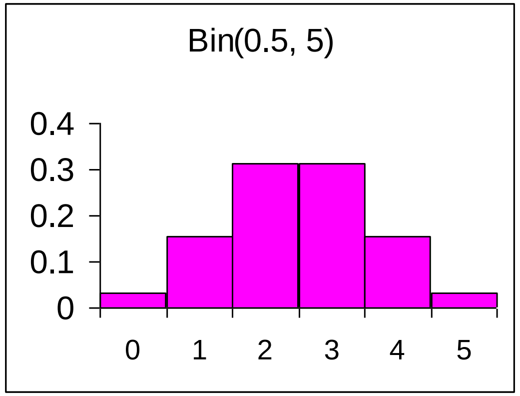
(b) more than 5 heads

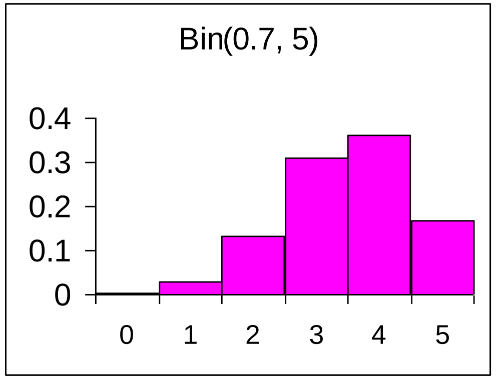
The Binomial Distribution

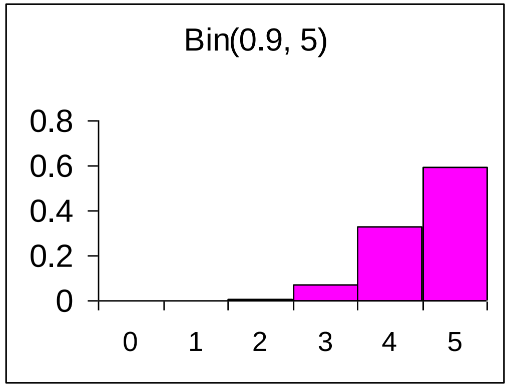
Overview











The Poisson Distribution
Overview

- When there is a large number of trials, but a small probability of success, binomial calculation becomes impractical
  - Example: Number of deaths from horse kicks in the Army in different years

- The mean number of successes from n trials is $\mu = np$
  - Example: 64 deaths in 20 years from thousands of soldiers

Simeon D. Poisson (1781-1840)

The Poisson Distribution
Overview

- If we substitute μ/n for p, and let n tend to infinity, the binomial distribution becomes the Poisson distribution:

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!}$$

The Poisson Distribution
Overview

- Poisson distribution is applied where random events in space or time are expected to occur
- Deviation from Poisson distribution may indicate some degree of non-randomness in the events under study
- Investigation of cause may be of interest

The Poisson Distribution

Question 3: In the manufacture of glassware, bubbles can occur in the glass which reduces the status of the glassware to that of a ‘second’. If, on average, one in every 1000 items produced has a bubble, calculate the probability that exactly six items in a batch of three thousand are seconds.

The Poisson Distribution

Suppose that:
X = number of items with bubbles, then X (3000, 0.001)

Since n = 3000 and p = 0.001, we can use the Poisson distribution with

$\mu = np = 3000 \times 0.001 = 3$

The calculation is:

$$P(X = 6) = \frac{e^{-3}3^6}{6!}$$

$$\approx 0.0498 \times 1.0125$$

$$\approx 0.05$$

The result means that we have about a 5% chance of finding exactly six seconds in a batch of three thousand items of glassware.

The Poisson Distribution

Question 4: A manufacturer produces light-bulbs that are packed into boxes of 100. If quality control studies indicate that 0.5% of the light-bulbs produced are defective, what percentage of the boxes will contain:

(a) no defective (b) less than 3 are defective

The Poisson Distribution
Emission of α-particles

- Rutherford, Geiger, and Bateman (1910) counted the number of α-particles emitted by a film of polonium in 2608 successive intervals of one-eighth of a minute
  - What is n?
  - What is p?

- Do their data follow a Poisson distribution?

The Poisson Distribution
Emission of α-particles

- Calculation of μ:

μ = No. of particles per interval

= 10097/2608

= 3.87

- Expected values:

$$2680 \times P(x) = 2608 \times \frac{e^{-3.87}(3.87)^x}{x!}$$

No. α-particles	Observed
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1
Over 14	0
Total	2608

The Poisson Distribution
Emission of α-particles

No. α-particles	Observed	Expected
0	57	54
1	203	210
2	383	407
3	525	525
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	0	1
13	1	1
14	1	1
Over 14	0	0
Total	2608	2680

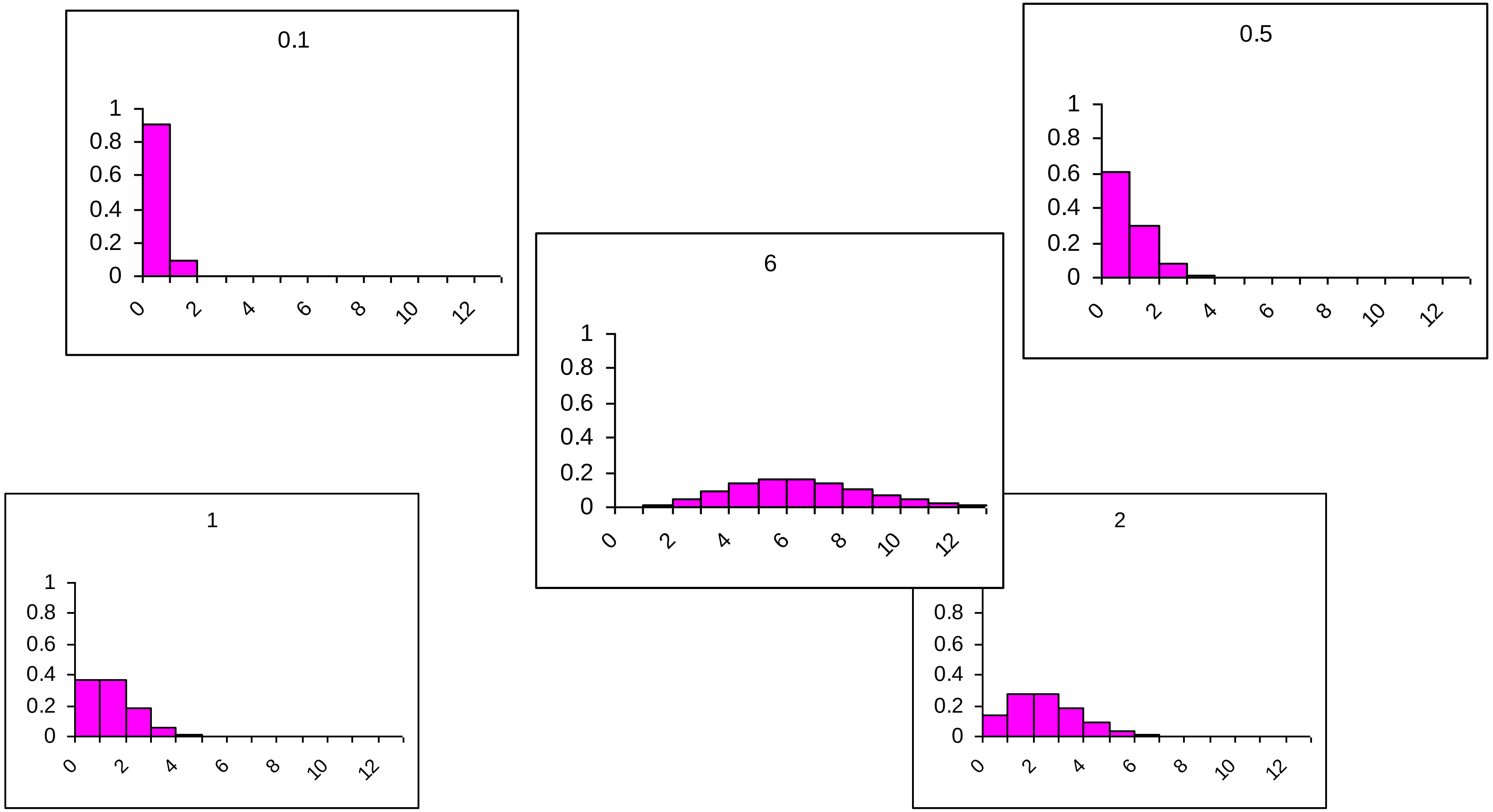
The Poisson Distribution
Emission of α-particles

Random events

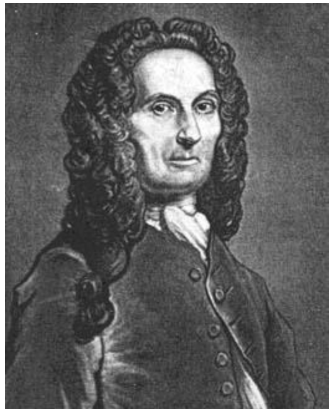
Regular events

Clumped events

The Poisson Distribution



The Normal Distribution
Overview



Abraham de Moivre
(1667-1754)

- Discovered in 1733 by de Moivre as an approximation to the binomial distribution when the number of trails is large
- Derived in 1809 by Gauss

- Importance lies in the Central Limit Theorem, which states that the sum of a large number of independent random variables (binomial, Poisson, etc.) will approximate a normal distribution
  - Example: Human height is determined by a large number of factors, both genetic and environmental, which are additive in their effects. Thus, it follows a normal distribution.

Karl F. Gauss
(1777-1855)

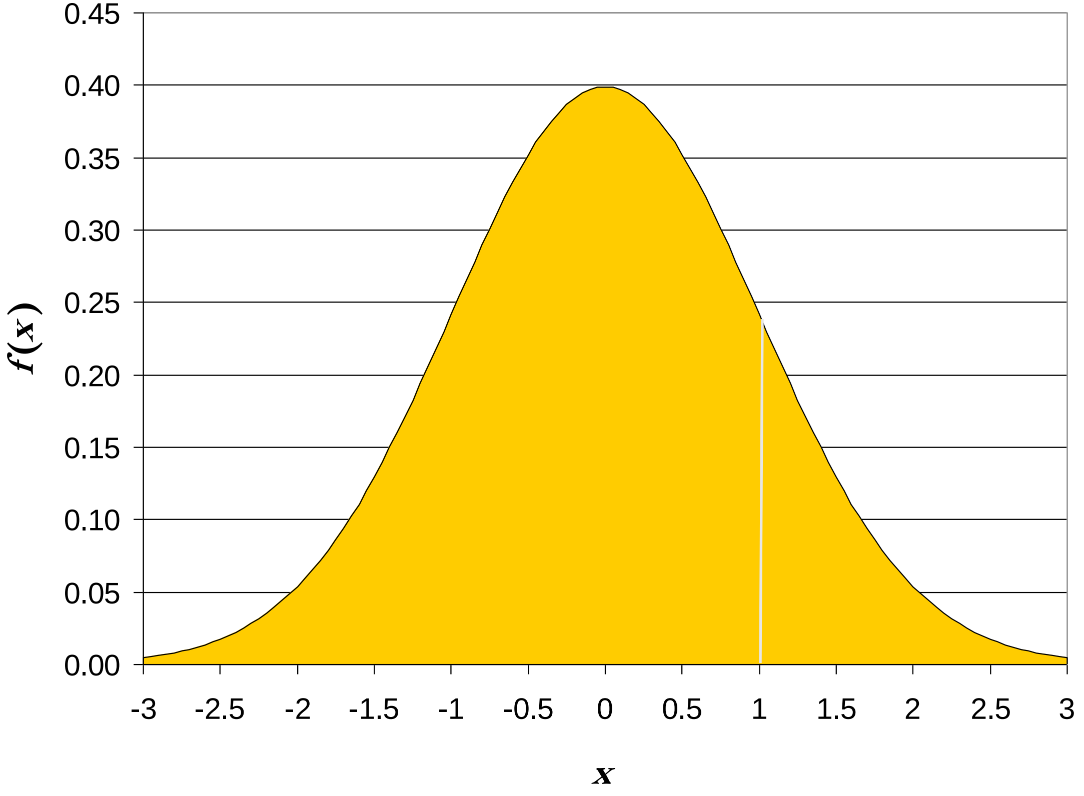
The Normal Distribution
Overview

- A <u>continuous</u> random variable is said to be normally distributed with mean μ and variance σ^2 if its probability <u>density</u> function is

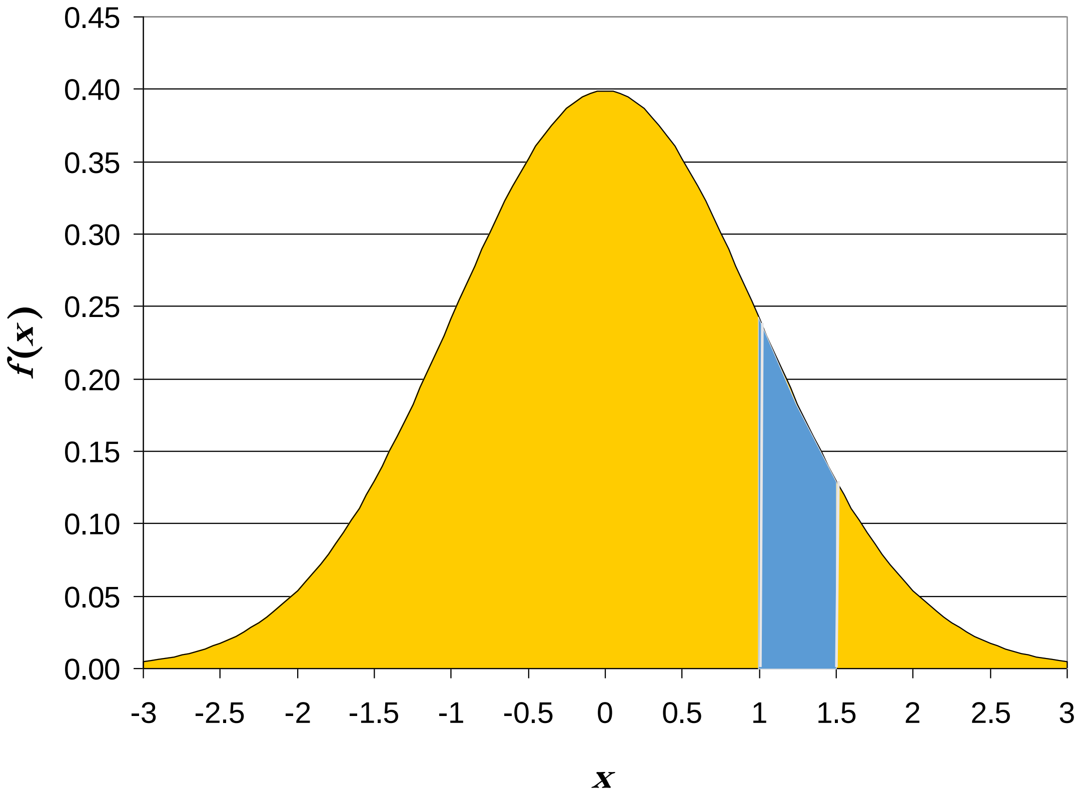
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

- $f(x)$ is not the same as $P(x)$
  - $P(x)$ would be 0 for every x because the normal distribution is continuous
  - However, $P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f(x)dx$

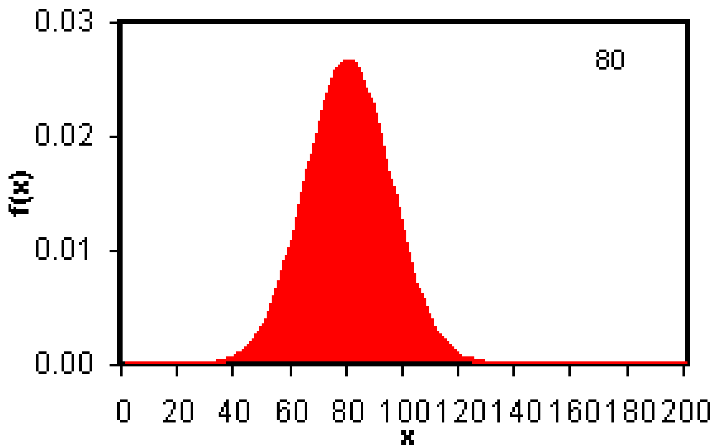
The Normal Distribution
Overview



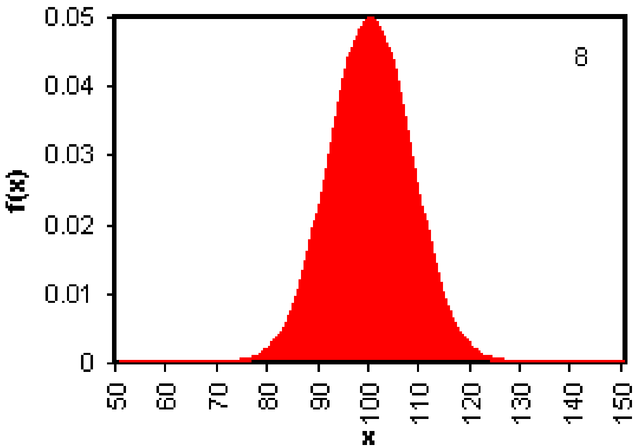
The Normal Distribution
Overview



The Normal Distribution
Overview



Mean changes



Variance changes

The Normal Distribution

Length of Fish

- A sample of rock cod in Monterey Bay suggests that the mean length of these fish is μ = 30 in. and σ^2 = 4 in.

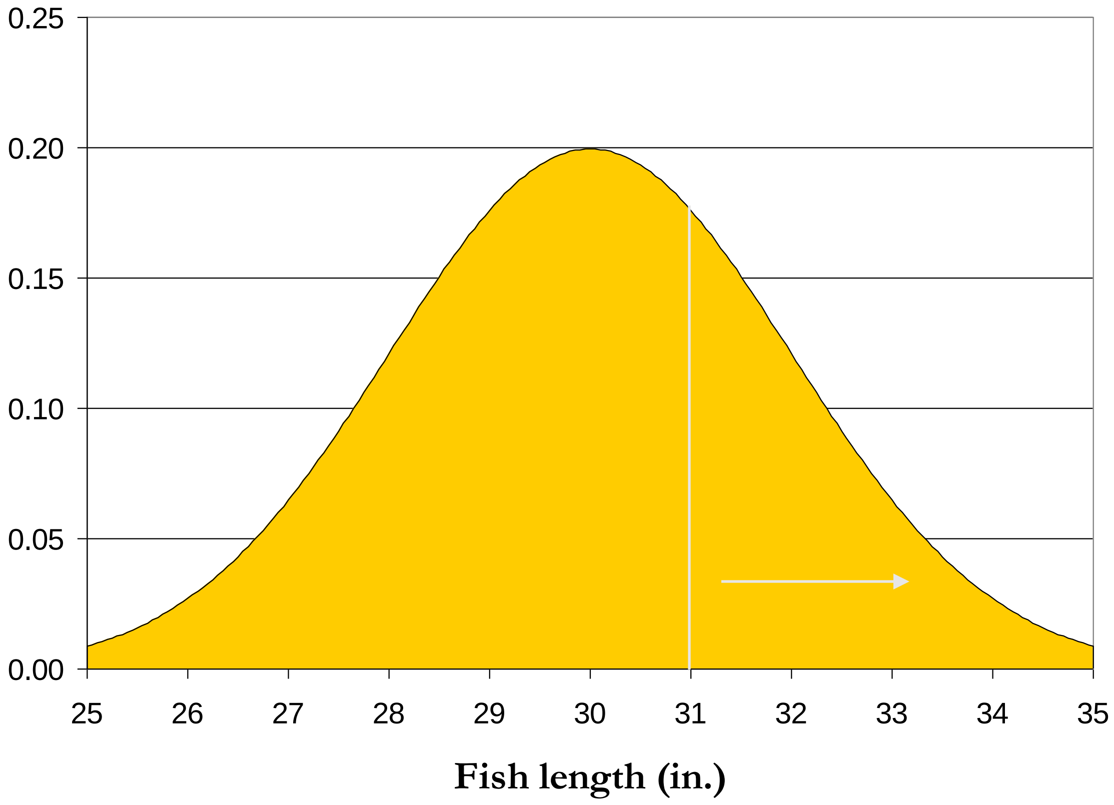
- Assume that the length of rock cod is a normal random variable

- If we catch one of these fish in Monterey Bay,
  - What is the probability that it will be at least 31 in. long?
  - That it will be no more than 32 in. long?
  - That its length will be between 26 and 29 inches?

The Normal Distribution

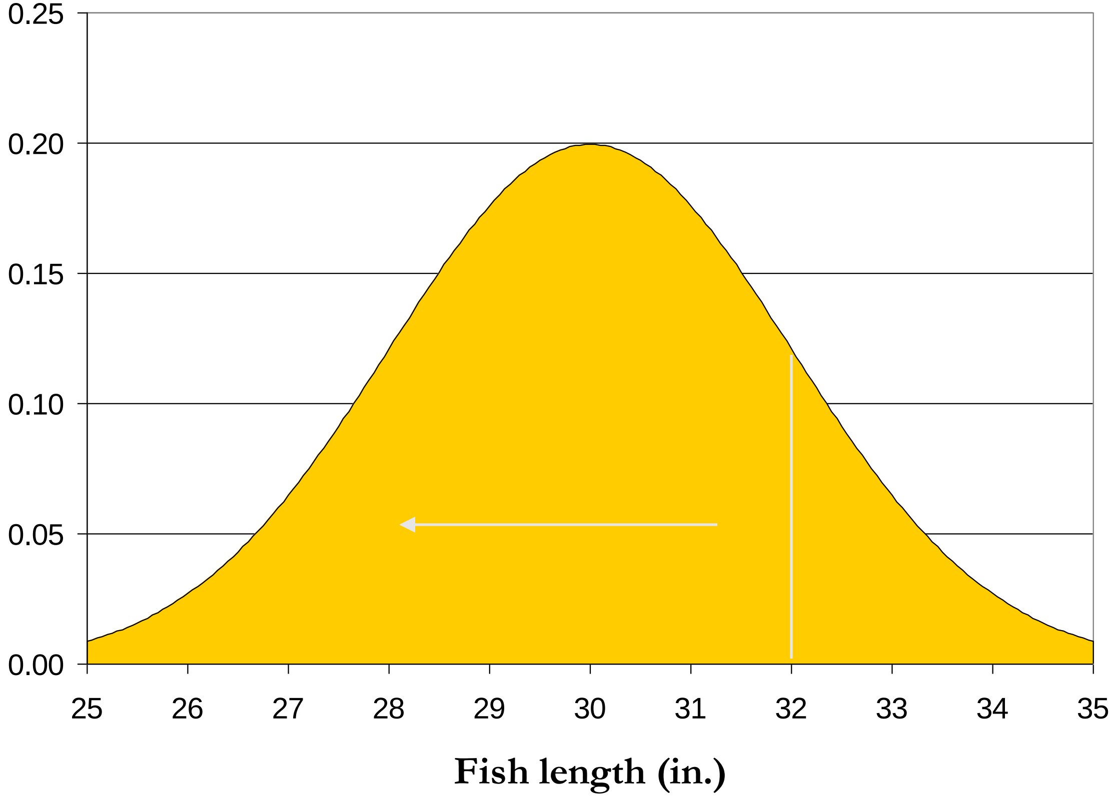
Length of Fish

- What is the probability that it will be at least 31 in. long?



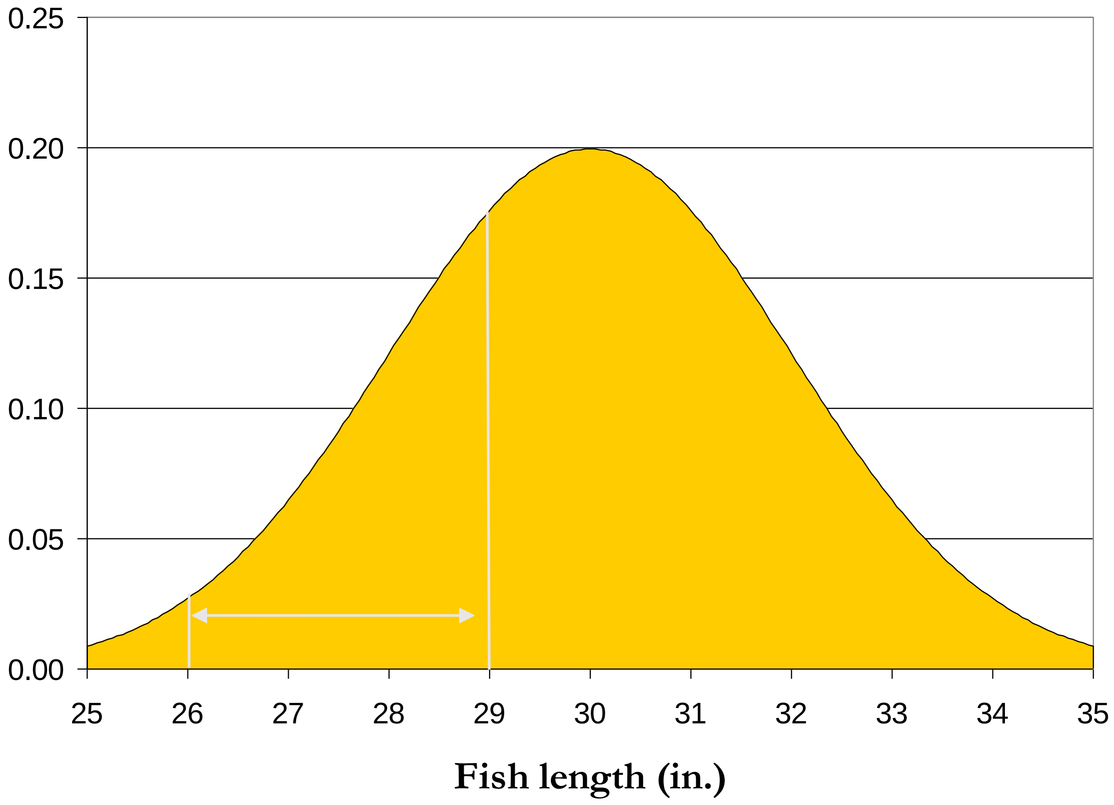
The Normal Distribution
Length of Fish

- That it will be no more than 32 in. long?



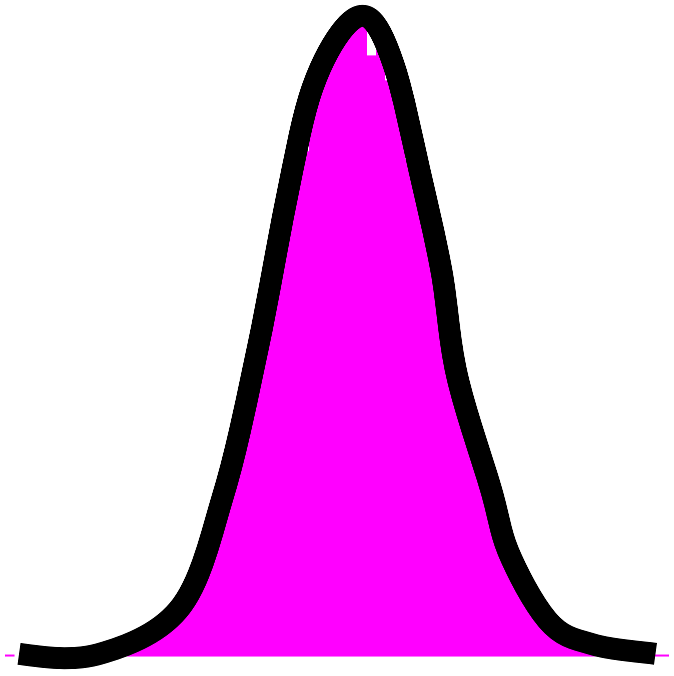
The Normal Distribution
Length of Fish

- That its length will be between 26 and 29 inches?



Standard Normal Distribution

- $\mu=0$ and $\sigma^2=1$



Thank you