Probability Distributions

S C Agarkar

Dr V N Bedekar Institute of Management Studies, Thane

Frequency and Probability Distribution

- Probability distributions are related to frequency distributions.
- We can think of probability distribution as a theoretical frequency distribution.
- Since these distributions deal with expectations they are useful tool in making inferences and decisions under conditions of uncertainity.

Random variable

- A random variable is a variable that takes different values as a result of the outcomes of a random experiment (examples, breast cancer).
- It is the mathematical rule (or function) that assigns a given numerical value to each possible outcome of an experiment in the sample space of interest.

Random variables

- Discrete random variables: It is allowed to take on only a limited number of values (e.g. data of a saree sale)
- Continuous random variables: It is allowed to assume any value within a given range (e.g. investments in the market).
- Expected value: It is the value that is guessed based on the data. Mathematically it is obtained by multiplying random variable by probability of occurrence and then summing these products.

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

- In general, if trials result in a series of success and failures,

  FFSFFFFSFSFSSFFFFFSF…

  Then the probability of x successes <u>in that order</u> is

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

The Binomial Distribution

- However, if order is not important, then

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

where $\frac{n!}{x!(n-x)!}$ is the number of ways to obtain x successes

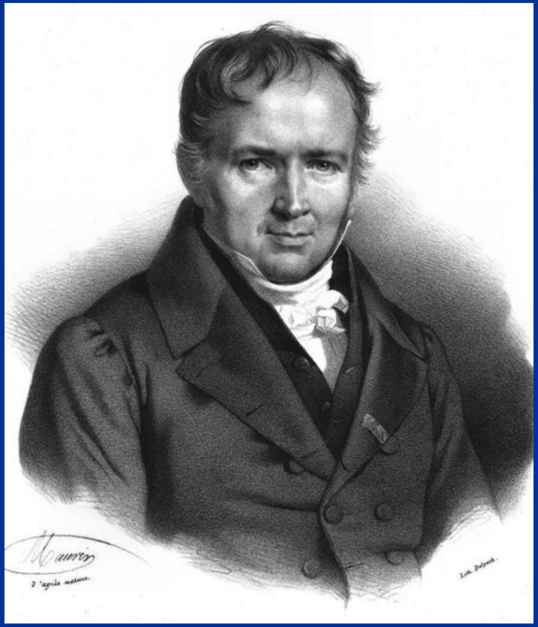
in n trials, and $i! = i \cdot (i-1) \cdot (i-2) \cdot \ldots \cdot 2 \cdot 1$

Example

- What is the probability of 2 successes in 3 trials in a coin tossing game.
- Probability of 2 successes in 3 trials
- $=3!/2!\ (3-2)\ (.5)^2\ (.5^1)$
- $= 6/2\ (.25)\ (.5)$
- $= 0.375$

The Poisson Distribution

- 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

- 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

- 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
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:

$$\mu = \text{No. of particles per interval}$$
$$= 10097/2608$$
$$= 3.87$$

- 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

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

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)

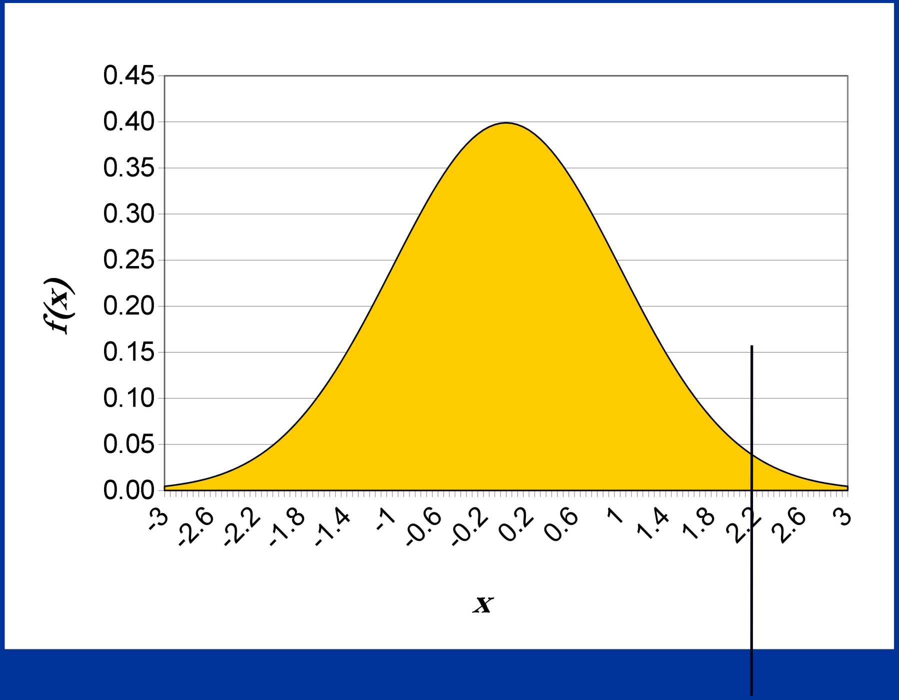
The Normal Distribution
Overview

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

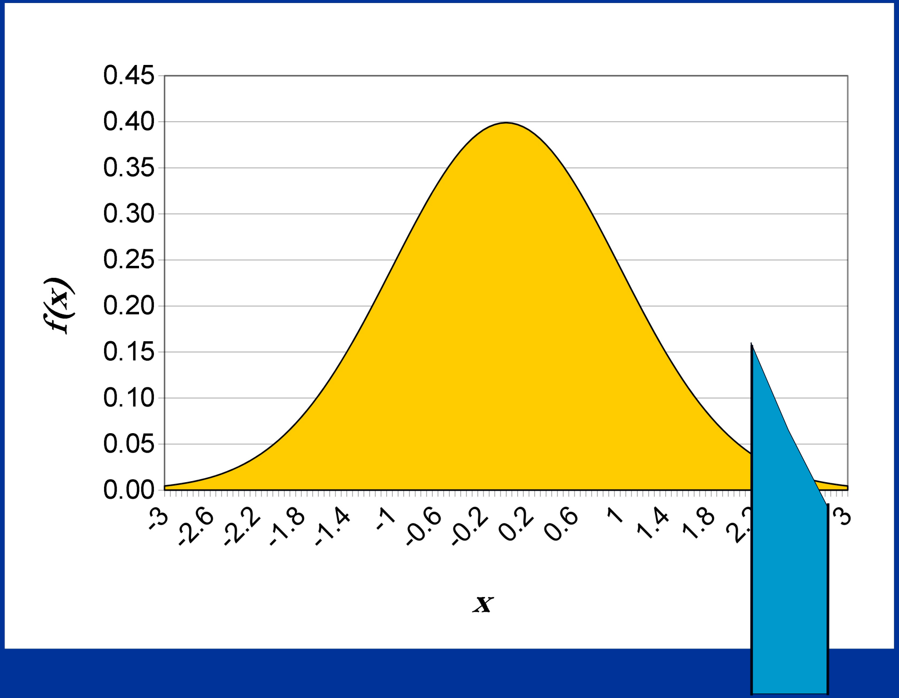
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

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

The Normal Distribution
Overview



The Normal Distribution
Overview

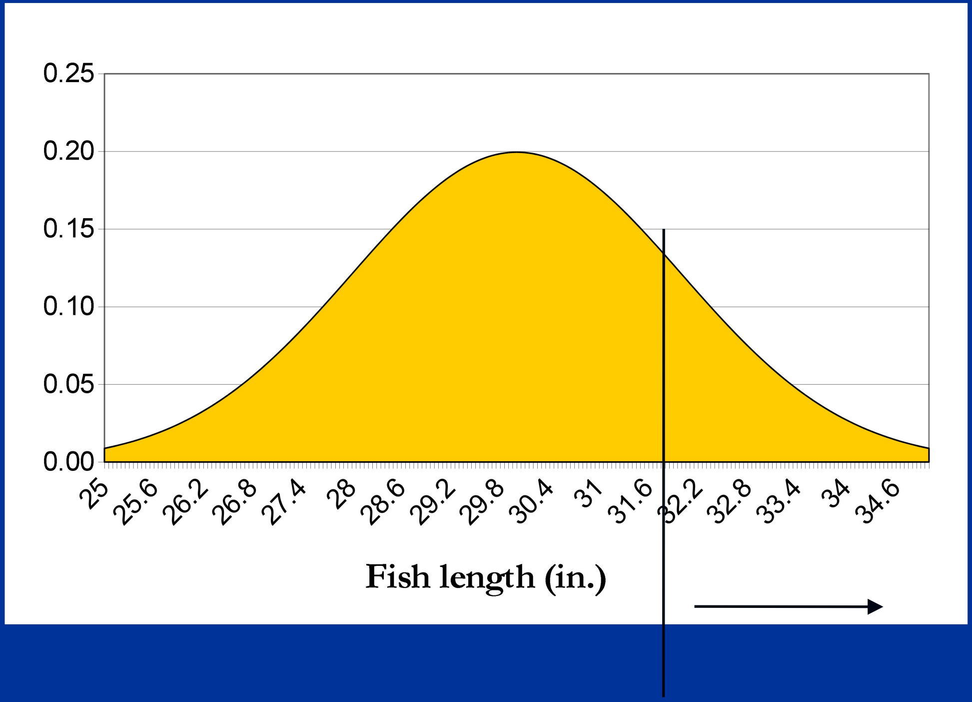


The Normal Distribution
Length of Fish

- A sample of rock cod in Monterey Bay suggests that the mean length of these fish is $\mu = 30$ in. and $\sigma^2 = 4$ 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?

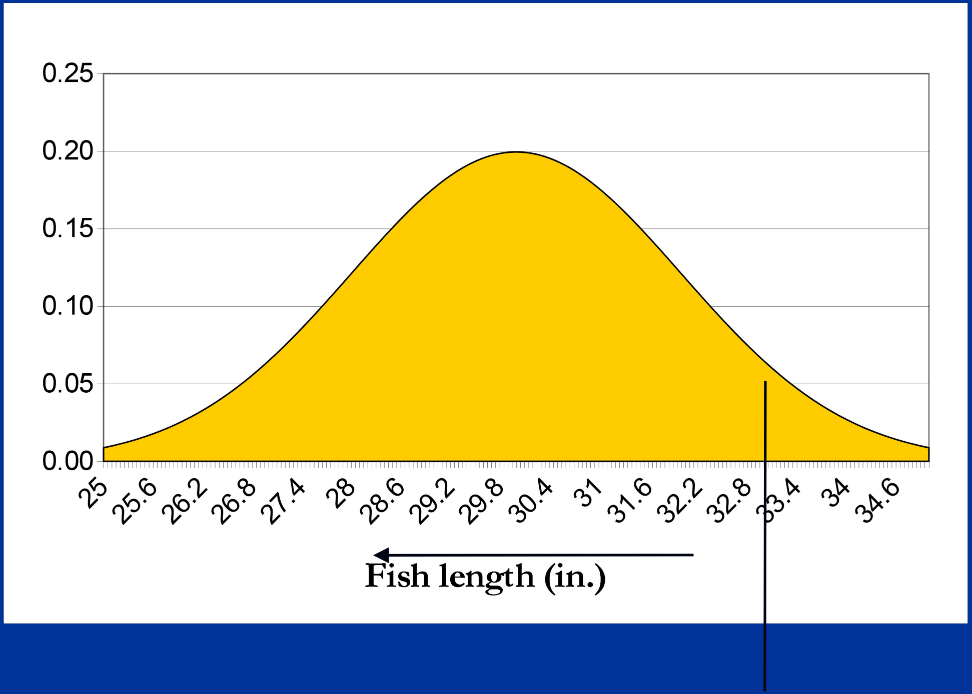
The Normal Distribution
Length of Fish

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



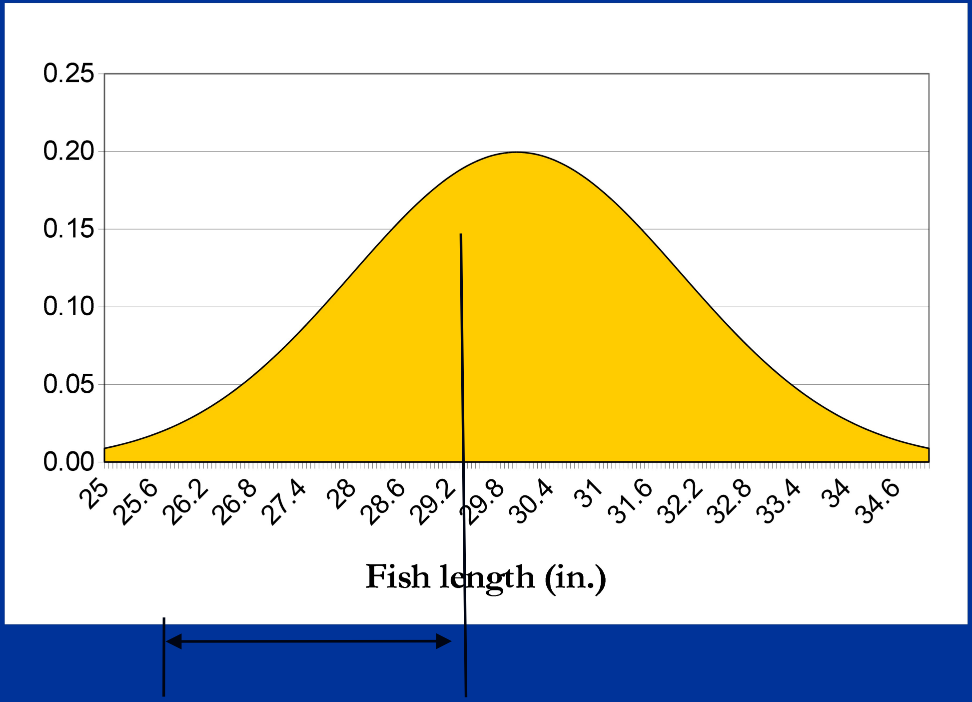
The Normal Distribution
Length of Fish

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



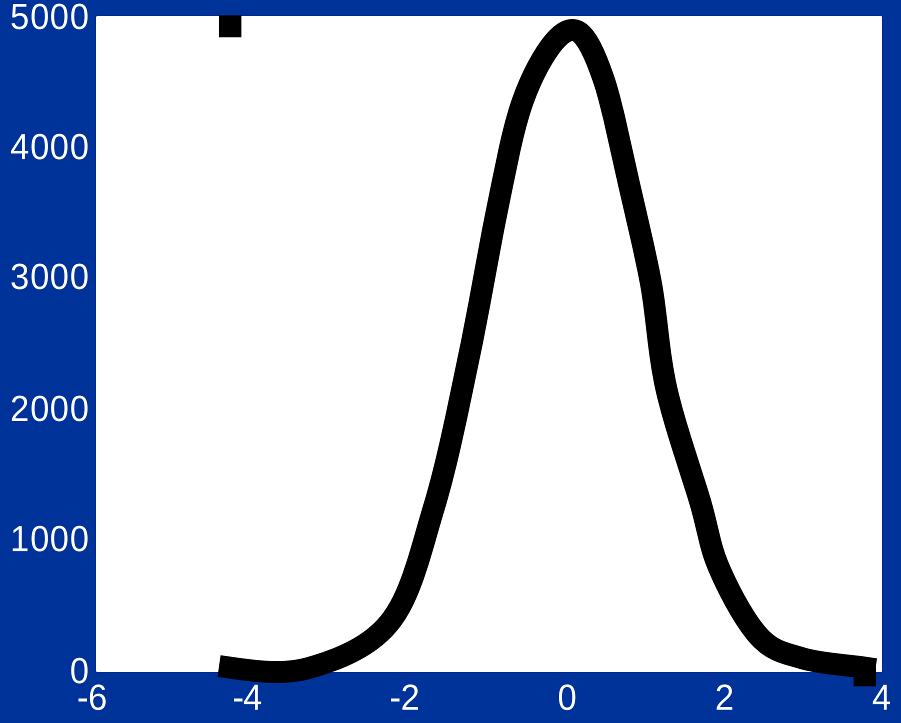
The Normal Distribution
Length of Fish

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



Standard Normal Distribution

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



Useful properties of the normal distribution

1. The normal distribution has useful properties:

- Can be added $E(X+Y)= E(X)+E(Y)$ and $\sigma 2(X+Y)= \sigma 2(X)+ \sigma 2(Y)$
- Can be transformed with ***shift*** and ***change of scale*** operations

Consider two random variables X anc

Let X~N(μ,σ) and let Y=aX+b where a and b area constants

Change of scale is the operation of multiplying *X* by a constant "***a***" because one unit of X becomes "a" units of Y.

Shift is the operation of adding a constant "*b*" to X because we simply move our random variable X "b" units along the x-axis.

If X is a normal random variable, then the new random variable Y created by this operations on X is also a random normal variable

For $X \sim N(\mu,\sigma)$ and $Y=aX+b$

- $E(Y) = a\mu+b$
- $\sigma^2(Y)=a^2\ \sigma^2$
- A special case of a change of scale and shift operation in which $a = 1/\sigma$ and $b = -1(\mu/\sigma)$
- $Y=(1/\sigma)X-\mu/\sigma$
- $Y=(X-\mu)/\sigma$ gives
- $E(Y)=0$ and $\sigma^2(Y) = 1$

The Central Limit Theorem

- That Standardizing any random variable that itself is a sum or average of a set of independent random variables results in a new random variable that is nearly the same as a standard normal one.
- The only caveats are that the sample size must be large enough and that the observations themselves must be independent and all drawn from a distribution with common expectation and variance.

Hypergeometric Distribution

- Hypergeometric distribution is a discrete probability distribution that describes the number of successes in a sequence of n draws from finite population without replacement.
- The basic characteristics of this distribution are
  - It is to be used when the population size is small.
  - The samples are drawn without replacement
  - The trials are independent.

Illustrative Example

- A class contains N students. Let M be the number of boys, that means girls would be (N-M), Now if we draw a sample of n students (without replacement) then the probability of getting k boys out of n students would be given as follows:
- $P(K) = [{}^{M}C_{k} * {}^{(N-M)}C_{n-k}] / {}^{N}C_{n}$
- This distribution is not useful in the situation when the population size is relatively large.

Example

- A box contains 30 items out of which 5 items are defective. What is the probability that if a sample of 8 is chosen at random
- 3 items will be defective (0.066)
- No item will be defective (0.184)
- At least one item will be defective (0.815)

Exponential Distribution

- When events occur continuously and independently at a constant average rate, the distribution followed by the random variable representing occurrence of event is said to follow Exponential distribution.
- Let $\lambda>0$ be a real number, the random variable X is said to be exponentially distributed if its probability density function is
- $f(x)= \lambda e^{-\lambda x}$

Example

- The average time the customer spends at a Pizza shop is 20 minutes. Find the probability that a customer has to spend more than 25 minutes at a shop.
- Given $\mu=20$, $\lambda=1/20$
- $Pr(X>25)= 1-Pr\ (X<25)$
- $= 0.286$

Uniform Distribution

- When equal probability is assigned to random variable for all the outcomes, it is the case of uniform distribution.
- Examples are tossing of coin or rolling of a dice as the probability of getting one result remains the same.
- The probability density of a uniformally distributed random variable x is given as
- $F(x)= 1/b-a$ if $a \leq x \leq b$

Example

- The time passenger shall wait at a ticket counter of a railway station is uniformally distributed on the interval .50. What is the probability that a passenger waits less than 15 minutes?
- F(x)= x-0/50-0=x/50
- Pr (X≤15)= f(15)
- = 15/50
- = 3/10

Thank You