Probability in Computing

LECTURE 8: CENTRAL LIMIT THEOREMS

Agenda

- Quick look at
  - Law of large numbers
  - Normal distribution
  - Central limit theorem

Law of large numbers

- The law of large numbers (LLN) is a theorem describes the result of performing the same experiment a large number of times:
  - The average of the results obtained from a large numb trials should be close to the expected value, and will te become closer as more trials are performed.
  - if a large number of dice are rolled, the average of the values (sometimes called the sample mean) is likely to close to 3.5, with the accuracy increasing as more dice rolled.

© 2010, Van Nguyen

LLN's importance

- The LLN is important because it "guarantees" st
long-term results for random events.
    - For example, while a casino may lose money in a singl of the roulette wheel, its earnings will tend towards a predictable percentage over a large number of spins.
    - Any winning streak by a player will eventually be overc by the parameters of the game. It is important to rem that the LLN only applies (as the name indicates) whe *large number* of observations are considered.

Basic idea of LLN

- With virtual certainty -- the sample average

$$\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

converges to the expected value

  - where X_1, X_2, ... is an infinite sequence of i.i.d. (indep and identically distributed) random variables with finit expected value

    $E(X_1) = E(X_2) = ... = \mu < \infty$.

- The strong law and the weak law:
  - The two versions are concerned with the mode of convergence being asserted.

Weak law of large numbers

- The sample average converges in probability tow the expected value

$$\overline{X}_n \xrightarrow{p} \mu \qquad \text{when } n \to \infty.$$

- That is to say that for any positive number *ε*,

$$\lim_{n\to\infty} \Pr\left(|\overline{X}_n - \mu| < \varepsilon \right) = 1.$$

  - the weak law essentially states that for any nonzero m specified, no matter how small, with a sufficiently larg sample there will be a very high probability that the av of the observations will be close to the expected value is, within the margin.

© 2010, Van Nguyen

Work in Class

- Create your own example that can illustrate this law
  - Hint: percentage of male population in districts of a country

Strong Law

- The strong law of large numbers states that sample average converges almost surely to the expected value

$$\overline{X}_n \xrightarrow{a.s.} \mu \qquad \text{when } n \to \infty.$$

- That is

$$\Pr\left(\lim_{n\to\infty} \overline{X}_n = \mu \right) = 1.$$

The normal distribution

- The normal distribution or Gaussian distribu
  - A continuous probability distribution that often gives a g description of data that cluster around the mean
  - The graph of the associated probability density function shaped, with a peak at the mean, and is known as the b curve.
  - The simplest case of a normal distribution is known as t standard normal distribution, described by pdf

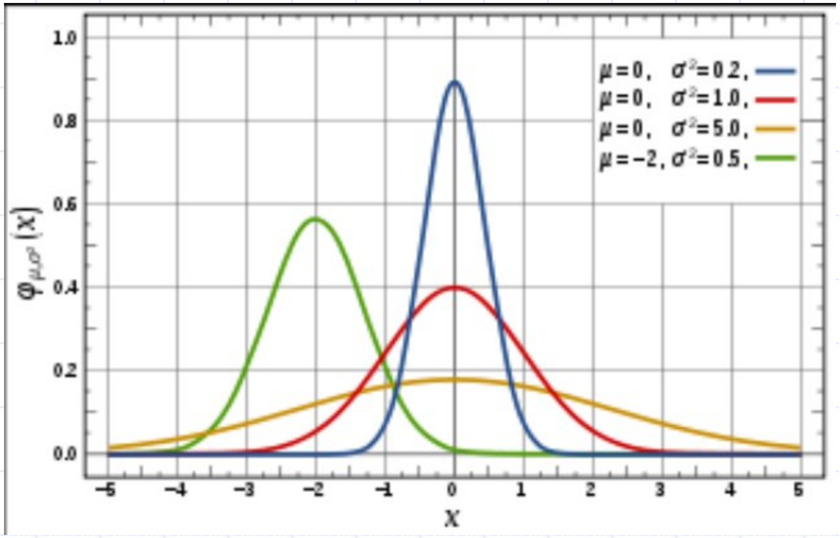
$$\phi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2},$$

  - More general with pdf $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\,\phi($
  - Thus when a random variable X is distributed normally w mean μ and variance σ^2, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

© 2010, Van Nguyen

The normal distribution

- The normal distribution or Gaussian distribu
is often used to describe, at least approximately,
variable that tends to cluster around the mean.
  - E.g., the heights of adult males in the United States are normally distributed, with a mean of about 70 inches (1 Most men have a height close to the mean, though a sn number of outliers have a height significantly above or b the mean.



© 2010, Van Nguyen

Central limit theorem

- By the central limit theorem, under certain conc the sum of a number of random variables with fi means and variances approaches a normal distri as the number of variables increases.
  - For this reason, the normal distribution is commonly encountered in practice, and is used throughout statist natural science, and social science as a simple model f complex phenomena.
  - For example, the observational error in an experiment assumed to follow a normal distribution
- The central limit theorem is also known as the s fundamental theorem of probability (first is LLN

Central limit theorem

- Let $X_1, X_2, X_3, \ldots, X_n$ be a sequence of n i.i.d. ran
variables each having finite values of expectation
variance $\sigma^2 > 0$. The CLT states that
  - as the sample size n increases the distribution of the sa
average of these random variables approaches the norn
distribution with a mean μ and variance σ^2/n irrespectiv
shape of the common distribution of the individual term
  - More precisely, let S_n be the sum $S_n = X_1 + \cdots + X_n$
  - Then, if we define new random variables $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$
  - then they will converge in distribution to the standard n
distribution $N(0,1)$ as n approaches infinity.

$$Z_n \xrightarrow{d} \mathcal{N}(0,1).$$

Work in Class

- Can you explain about the normal distribution of heights of adult men i
country?