

UCLA STAT 35
Applied Computational and Interactive Probability

• **Instructor:** Ivo Dinov,
Asst. Prof. In Statistics and Neurology

• **Teaching Assistant:** Chris Barr

University of California, Los Angeles, Winter 2006
<http://www.stat.ucla.edu/~dinov/>

Stat 35, UCLA, Ivo Dinov Slide 1

**Statistics & Their Distributions –
The CLT**

Slide 2 Stat 35, UCLA, Ivo Dinov

Statistic

A *statistic* is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. A statistic is a random variable denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

Slide 3 Stat 35, UCLA, Ivo Dinov

Random Samples

The rv's X_1, \dots, X_n are said to form a (simple *random sample* of size n if

1. The X_i 's are independent rv's.
2. Every X_i has the same probability distribution.

Slide 4 Stat 35, UCLA, Ivo Dinov

Simulation Experiments

The following characteristics must be specified:

1. The statistic of interest.
2. The population distribution.
3. The sample size n .
4. The number of replications k .

Slide 5 Stat 35, UCLA, Ivo Dinov

The Distribution of the Sample Mean

Slide 6 Stat 35, UCLA, Ivo Dinov

Using the Sample Mean

Let X_1, \dots, X_n be a random sample from a distribution with mean value μ and standard deviation σ . Then

$$1. E(\bar{X}) = \mu_{\bar{X}} = \mu$$

$$2. V(\bar{X}) = \sigma_{\bar{X}}^2 = \sigma^2/n$$

In addition, with $T_o = X_1 + \dots + X_n$,
 $E(T_o) = n\mu$, $V(T_o) = n\sigma^2$, and $\sigma_{T_o} = \sqrt{n}\sigma$.

Slide 7 Stat 35, UCLA, Ivo Dinov

Normal Population Distribution

Let X_1, \dots, X_n be a random sample from a normal distribution with mean value μ and standard deviation σ . Then for any n , \bar{X} is normally distributed, as is T_o .

<http://socr.stat.ucla.edu/Applets.dir/SamplingDistributionApplet.html>

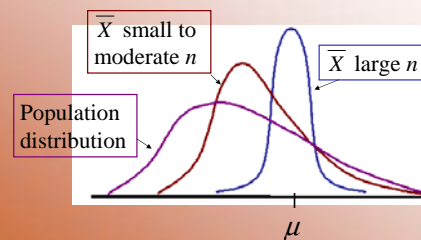
Slide 8 Stat 35, UCLA, Ivo Dinov

The Central Limit Theorem

Let X_1, \dots, X_n be a random sample from a distribution with mean value μ and variance σ^2 . Then if n sufficiently large, \bar{X} has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$, and T_o also has approximately a normal distribution with $\mu_{T_o} = n\mu$, $\sigma_{T_o}^2 = n\sigma^2$. The larger the value of n , the better the approximation.

Slide 9 Stat 35, UCLA, Ivo Dinov

The Central Limit Theorem



Slide 10 Stat 35, UCLA, Ivo Dinov

Rule of Thumb

If $n > 30$, the Central Limit Theorem can be used.

Slide 11 Stat 35, UCLA, Ivo Dinov

Approximate Lognormal Distribution

Let X_1, \dots, X_n be a random sample from a distribution for which only positive values are possible [$P(X_i > 0) = 1$]. Then if n is sufficiently large, the product $Y = X_1 X_2 \dots X_n$ has approximately a lognormal distribution.

Slide 12 Stat 35, UCLA, Ivo Dinov

Central Limit Theorem – heuristic formulation

Central Limit Theorem:

When sampling from almost any distribution,

\bar{X} is approximately **Normally distributed** in **large samples**.

Show Sampling Distribution Simulation Applet:
 file:///C:/Ivo.dir/UCLA_Classes/Winter2002/AdditionalInstructorAids/
[SamplingDistributionApplet.html](#)

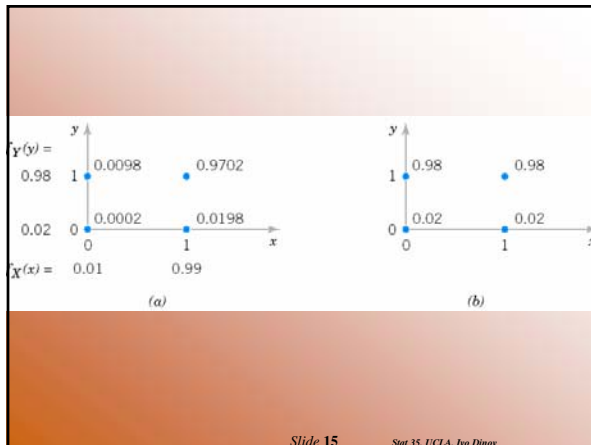
Slide 13 Stat 35, UCLA, Ivo Dinov

Independence

- For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are independent.

- $f_{XY}(x,y) = f_X(x) f_Y(y)$ for all x and y
- $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$
- $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$
- $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y respectively.

Slide 14 Stat 35, UCLA, Ivo Dinov



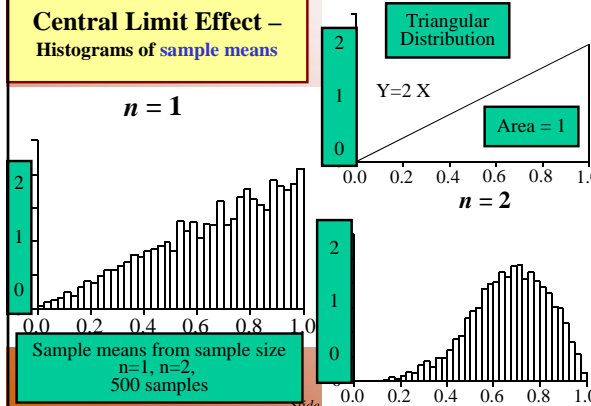
Slide 15 Stat 35, UCLA, Ivo Dinov

Recall we looked at the sampling distribution of \bar{X}

- For the sample mean calculated from a random sample, $E(\bar{X}) = \mu$ and $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, provided $\bar{X} = (X_1 + X_2 + \dots + X_n)/n$, and $X_k \sim N(\mu, \sigma)$. Then $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. And variability from sample to sample in the **sample-means** is given by the variability of the individual observations divided by the square root of the sample-size. In a way, averaging decreases variability.

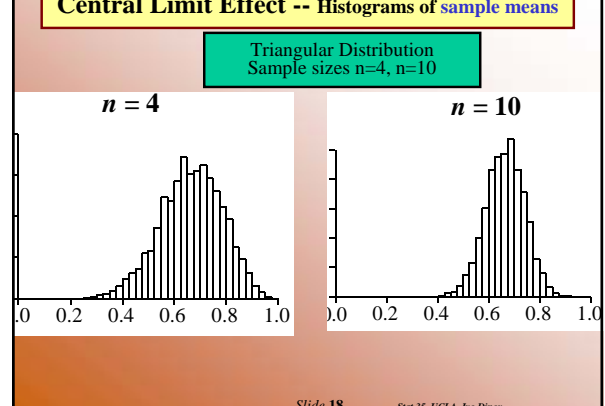
Slide 16 Stat 35, UCLA, Ivo Dinov

Central Limit Effect – Histograms of sample means

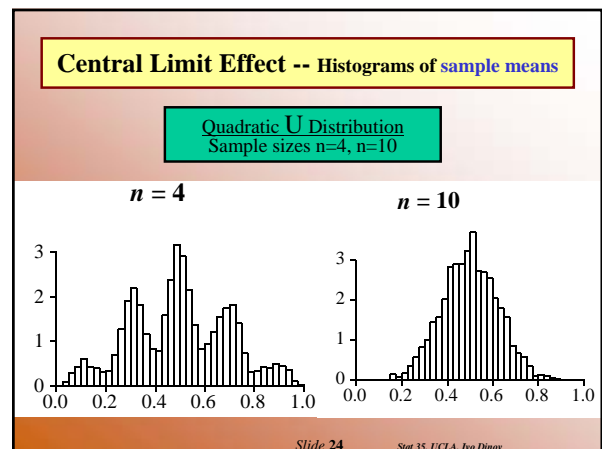
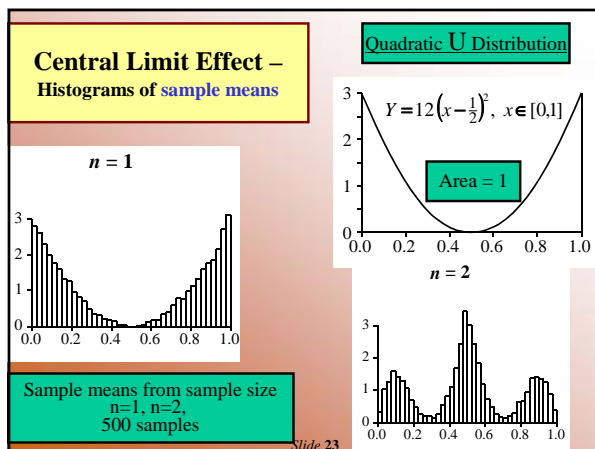
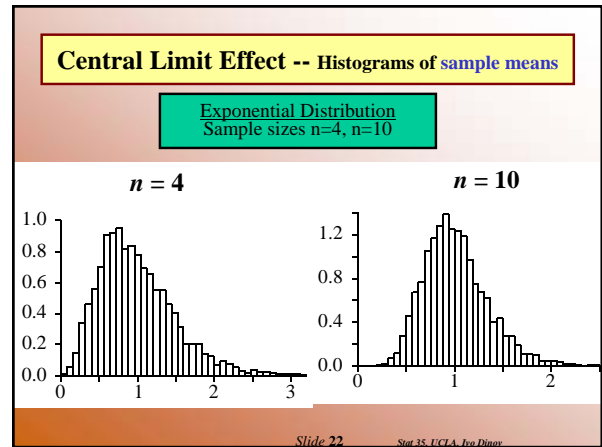
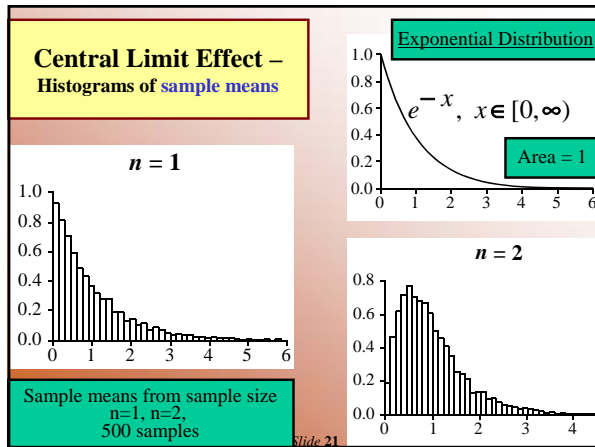
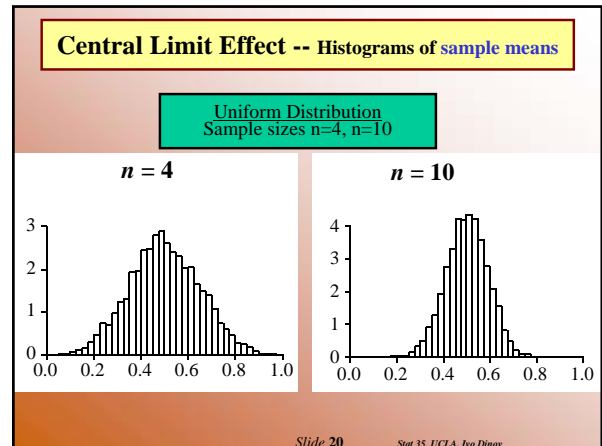
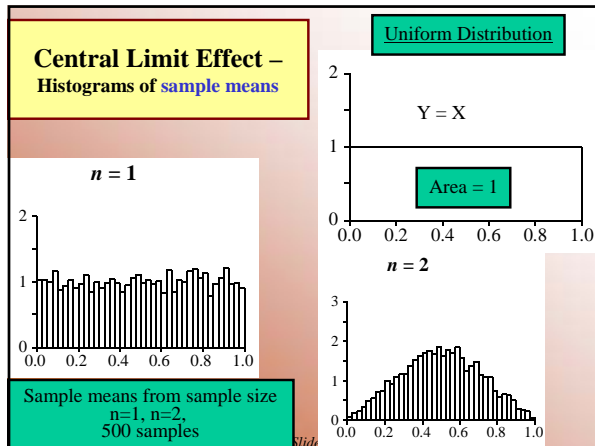


Slide 17

Central Limit Effect -- Histograms of sample means



Slide 18 Stat 35, UCLA, Ivo Dinov



Central Limit Theorem – heuristic formulation

Central Limit Theorem:

When sampling from almost any distribution,

\bar{X} is approximately Normally distributed in large samples.

Show Sampling Distribution Simulation Applet:
file://C:/Ivo.dir/UCLA_Classes/Winter2002/AdditionalInstructorAids/
[SamplingDistributionApplet.html](#)

Slide 25 Stat 35, UCLA, Ivo Dinov

Central Limit Theorem – theoretical formulation

Let $\{X_1, X_2, \dots, X_k, \dots\}$ be a sequence of independent observations from one specific random process. Let and $E(X) = \mu$ and $SD(X) = \sigma$ and both be finite ($0 < \sigma < \infty$; $|\mu| < \infty$). If $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, sample-avg,

Then \bar{X} has a distribution which approaches $N(\mu, \sigma^2/n)$, as $n \rightarrow \infty$.

Slide 26 Stat 35, UCLA, Ivo Dinov

The Distribution of a Linear Combination

Slide 27 Stat 35, UCLA, Ivo Dinov

Linear Combination

Given a collection of n random variables X_1, \dots, X_n and n numerical constants a_1, \dots, a_n , the rv

$$Y = a_1 X_1 + \dots + a_n X_n = \sum_{i=1}^n a_i X_i$$

is called a *linear combination* of the X_i 's.

Slide 28 Stat 35, UCLA, Ivo Dinov

Expected Value of a Linear Combination

Let X_1, \dots, X_n have mean values $\mu_1, \mu_2, \dots, \mu_n$ and variances of $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, respectively

Whether or not the X_i 's are independent,

$$E(a_1 X_1 + \dots + a_n X_n) = a_1 E(X_1) + \dots + a_n E(X_n) \\ = a_1 \mu_1 + \dots + a_n \mu_n$$

Slide 29 Stat 35, UCLA, Ivo Dinov

Variance of a Linear Combination

If X_1, \dots, X_n are independent,

$$V(a_1 X_1 + \dots + a_n X_n) = a_1^2 V(X_1) + \dots + a_n^2 V(X_n) \\ = a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2$$

and

$$\sigma_{a_1 X_1 + \dots + a_n X_n} = \sqrt{a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2}$$

Slide 30 Stat 35, UCLA, Ivo Dinov

Variance of a Linear Combination

For any X_1, \dots, X_n ,

$$V(a_1X_1 + \dots + a_nX_n) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)$$

Slide 31 Stat 35, UCLA, Jon Dineen

Difference Between Two Random Variables

$$E(X_1 - X_2) = E(X_1) - E(X_2)$$

and, if X_1 and X_2 are independent,

$$V(X_1 - X_2) = V(X_1) + V(X_2)$$

Slide 32 Stat 35, UCLA, Jon Dineen

Difference Between Normal Random Variables

If X_1, X_2, \dots, X_n are independent, normally distributed rv's, then any linear combination of the X_i 's also has a normal distribution. The difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.

Slide 33 Stat 35, UCLA, Jon Dineen