UCLA STAT 110 A

Applied Probability & Statistics for Engineers

•**Instructor**: **Ivo Dinov,**

Asst. Prof. In Statistics and Neurology

•**Teaching Assistant:** **Maria Chang, UCLA Statistics**

University of California, Los Angeles, Spring 2003

http://www.stat.ucla.edu/~dinov/

Chapter 5: Marginal & Joint PDF's Central Limit Theorem (CLT)

- Jointly distributed RV's
- Expected Value, Covariance, Correlation
- Distributions of sample statistics
- Central Limit Theorem (CLT)

Introduction

- If X and Y are two random variables, the probability distribution that defines their simultaneous behavior is a ***Joint Probability Distribution.***

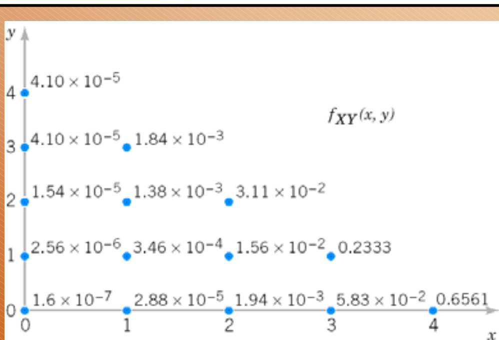
Examples:

- Signal transmission: X is high quality signals and Y low quality signals.
- Molding: X is the length of one dimension of molded part, Y is the length of another dimension.

- THUS, we may be interested in expressing probabilities expressed in terms of X and Y, e.g., ***P(2.95<X<3.05 and 7.60<Y<7.8)***

Two discrete random variables

- ***Range of random variables (X,Y)*** is the set of points (x,y) in 2D space for which the probability that X=x and Y=y is positive.
- If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x,y) in the range of (X,Y) along with the probability of each point.
- Sometimes referred to as ***Bivariate probability distribution***, or ***Bivariate distribution.***



Joint probability distribution

Joint probability mass function

- The joint probability mass function of the discrete random variables X and Y, denoted as $f_{XY}(x,y)$ satisfies:

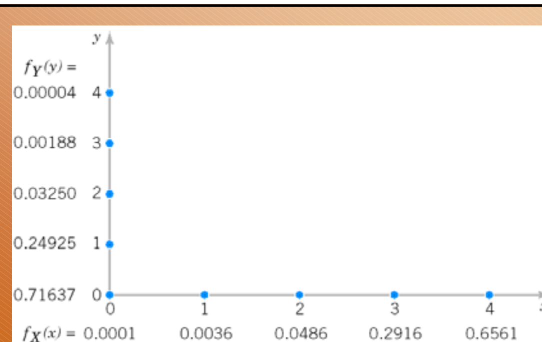
$$(1)\, f_{XY}(x,y) \geq 0$$

$$(2)\sum_x \sum_y f_{XY}(x,y) = 1$$

$$(3)\, f_{XY}(x,y) = P(X = x, Y = y)$$

Marginal probability distributions

- Individual probability distribution of a random variable is referred to as its ***Marginal Probability Distribution.***
- Marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables.
- Example 5-3: ***Marginal probability distribution of X is found by summing the probabilities in each column***, ***for y, summation is done in each row.***



Marginal probability distribution for X and Y

Marginal probability distributions (Cont.)

- If X and Y are discrete random variables with joint probability mass function $f_{XY}(x,y)$, then the marginal probability mass function of X and Y are

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(X,Y)$$

$$f_Y(y) = P(Y = y) = \sum_{Ry} f_{XY}(X,Y)$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and Ry denotes the set of all points in the range of (X, Y) for which $Y = y$

Mean and Variance

- If the marginal probability distribution of X has the probability function f(x), then

$$E(X) = \mu_X = \sum_x x f_X(x) = \sum_x x\left(\sum_{R_x} f_{XY}(x,y)\right) = \sum_x \sum_{R_x} x f_{XY}(x,y)$$
$$= \sum_R x f_{XY}(x,y)$$

$$V(X) = \sigma^2{}_X = \sum_x (x-\mu_X)^2 f_X(x) = \sum_x (x-\mu_X)^2 \sum_{R_x} f_{XY}(x,y)$$
$$= \sum_x \sum_{R_x} (x-\mu_X)^2 f_{XY}(x,y) = \sum_R (x-\mu_X)^2 f_{XY}(x,y)$$

- R = Set of all points in the range of (X,Y).
- Example 5-4.

Joint probability mass function – example

The joint density, P{*X,Y*}, of the number of minutes waiting to catch the first fish, *X* , and the number of minutes waiting to catch the second fish, *Y*, is given below.

P {*X* = i, *Y* = k }	k 1	2	3	Row Sum P{ *X* = i }
i 1	0.01	0.02	0.08	0.11
i 2	0.01	0.02	0.08	0.11
i 3	0.07	0.08	0.63	0.78
Column Sum P {*Y* =k }	0.09	0.12	0.79	1.00

- The (joint) chance of **waiting 3 minutes to catch the first fish and 3 minutes to catch the second fish** is:
- The (marginal) chance of **waiting 3 minutes to catch the first fish** is:
- The (marginal) chance of **waiting 2 minutes to catch the first fish** is (circle all that are correct):
- The chance of **waiting at least two minutes to catch the first fish** is (circle none, one or more):
- The chance of **waiting at most two minutes to catch the first fish and at most two minutes to catch the second fish** is (circle none, one or more):

Conditional probability

- Given discrete random variables X and Y with joint probability mass function $f_{XY}(X,Y)$, the conditional probability mass function of Y given X=x is

$f_{Y|x}(y|x) = f_{Y|x}(y) = f_{XY}(x,y)/f_X(x)$ for $f_X(x) > 0$

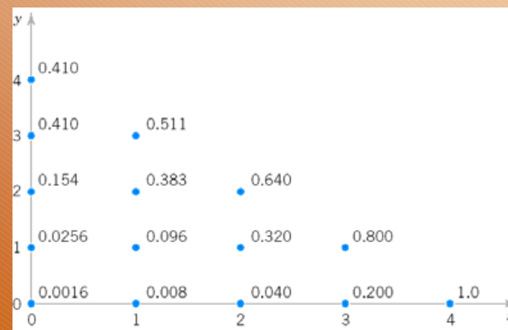
Conditional probability (Cont.)

- Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \ge 0$

(2) $\sum_{R_x} f_{Y|x}(y) = 1$

(3) $P(Y=y|X=x) = f_{Y|x}(y)$



Conditional probability (Cont.)

- Let R_x denote the set of all points in the range of (X,Y) for which X=x. The conditional mean of Y given X=x, denoted as E(Y|x) or $\mu_{Y|x}$, is

$$E(Y \mid x) = \sum_{R_x} y f_{Y|x}(y)$$

- And the conditional variance of Y given X=x, denoted as V(Y|x) or $\sigma^2_{Y|x}$ is

$$V(Y \mid x) = \sum_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_{R_x} y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

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.

(1) $f_{XY}(x,y) = f_X(x)\, f_Y(y)$ for all x and y

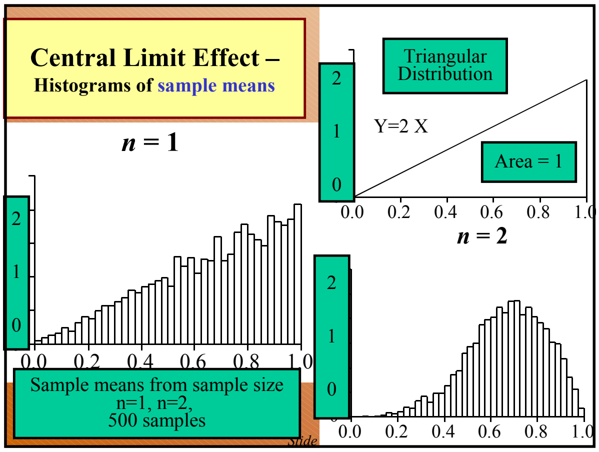
(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

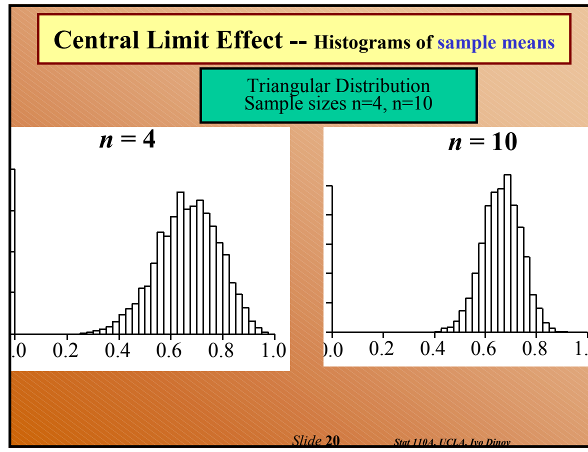
(3) $f_{X|y}(y) = f_X(x)$ for all x and y with $f_Y(y) > 0$

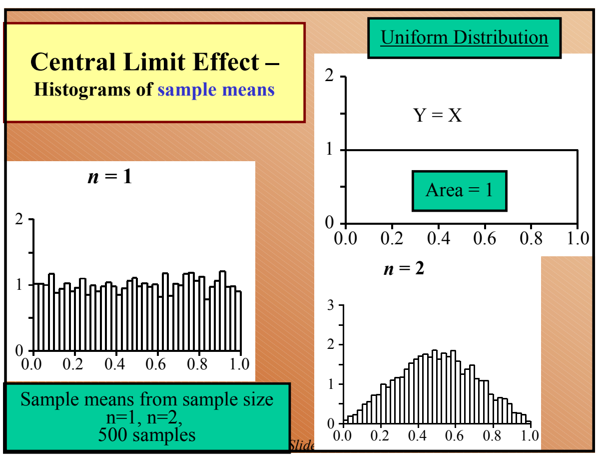
(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y respectively.

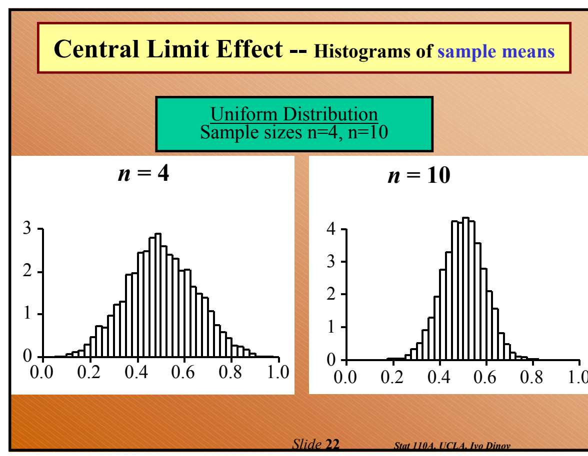
Recall we looked at the sampling distribution of $\bar{X}$

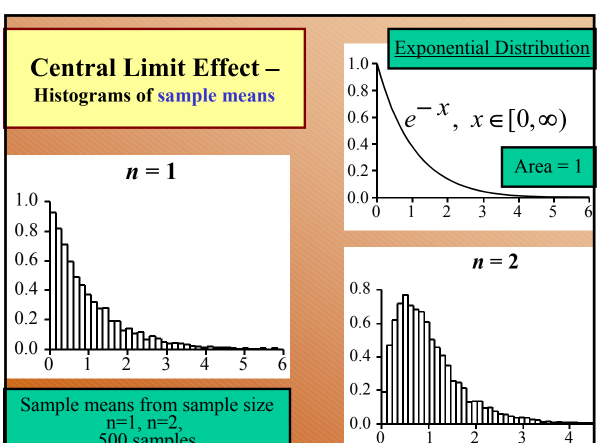
- For the sample mean calculated from a random sample, $E(\bar{X}) = \mu$ and $SD(\bar{X}) = \sigma/\sqrt{n}$, provided $\bar{X} = (X_1+X_2+ \ldots + X_n)/n$, and $X_k \sim N(\mu, \sigma)$. Then
- $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. And variability from sample to sample in the ***sample-means***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

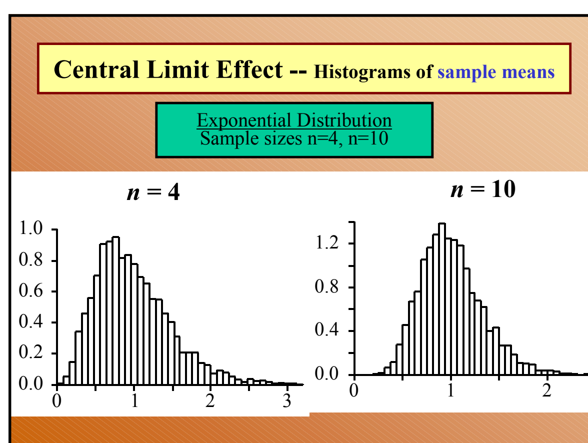

Central Limit Effect –
Histograms of sample means
n = 1
Triangular Distribution
Y=2 X
Area = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples


Central Limit Effect -- Histograms of sample means
Triangular Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 20
Stat 110A, UCLA, Ivo Dinov


Central Limit Effect –
Histograms of sample means
Uniform Distribution
Y = X
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples

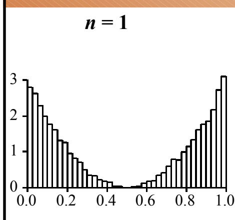

Central Limit Effect -- Histograms of sample means
Uniform Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 22
Stat 110A, UCLA, Ivo Dinov

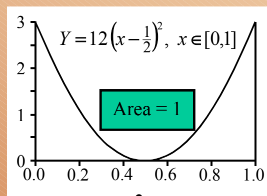

Central Limit Effect –
Histograms of sample means
Exponential Distribution
$e^{-x}, \ x \in [0, \infty)$
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples
Slide 23

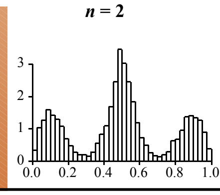

Central Limit Effect -- Histograms of sample means
Exponential Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 24
Stat 110A, UCLA, Ivo Dinov

Central Limit Effect – Histograms of sample means

Quadratic U Distribution



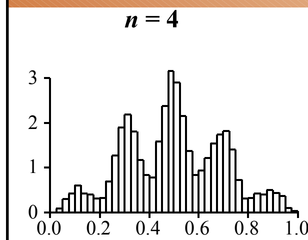


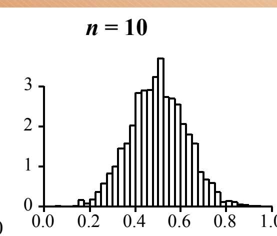


Sample means from sample size n=1, n=2, 500 samples

Central Limit Effect -- Histograms of sample means

Quadratic U Distribution
Sample sizes n=4, n=10





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

Central Limit Theorem – theoretical formulation

Let $\{X_1, X_2, \ldots, X_k, \ldots\}$ 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 $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, sample-avg,

Then $\bar{X}$ has a distribution which approaches N(μ, σ²/n), as $n \to \infty$.

Review

- What does the central limit theorem say? Why is it useful? (If the sample sizes are large, the mean in Normally distributed, as a RV)
- In what way might you expect the central limit effect to differ between samples from a symmetric distribution and samples from a very skewed distribution? (Larger samples for non-symmetric distributions to see CLT effects)
- What other important factor, apart from skewness, slows down the action of the central limit effect?

(Heavyness in the tails of the original distribution.)

Review

- When you have data from a moderate to small sample and want to use a normal approximation to the distribution of $\bar{X}$ in a calculation, what would you want to do before having any faith in the results? (30 or more for the sample-size, depending on the skewness of the distribution of *X*. Plot the data - non-symmetry and heavyness in the tails slows down the CLT effects).
- Take-home message: CLT is an application of statistics of paramount importance. Often, we are not sure of the distribution of an observable process. However, the CLT gives us a theoretical description of the distribution of the sample means as the sample-size increases (N(μ, σ²/n)).

The standard error of the mean – remember …

- For the sample mean calculated from a random sample, SD($\overline{X}$) = $\frac{\sigma}{\sqrt{n}}$. This implies that the 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.
- Recall that for *known* SD(X)=σ, we can express the SD($\overline{X}$) = $\frac{\sigma}{\sqrt{n}}$. How about if SD(X) is *unknown*?!?

The standard error of the mean

The standard error of the sample mean is an estimate of the *SD* of the sample mean

- i.e. a measure of the precision of the sample mean as an estimate of the population mean
- given by $SE(\overline{x}) = \frac{\text{Sample standard deviation}}{\sqrt{\text{Sample size}}}$

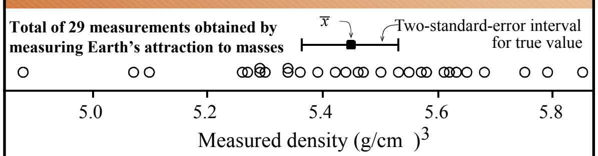
$$SE(\overline{x}) = \frac{s_x}{\sqrt{n}}$$

- Note similarity with
- SD($\overline{X}$) = $\frac{\sigma}{\sqrt{n}}$.

Cavendish's 1798 data on mean density of the Earth, g/cm^3, relative to that of H_2O

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

Source: Cavendish [1798].



Newton's law of gravitation: $F = G\, m_1\, m_2 / r^2$, the attraction force F is the ratio of the product (Gravitational const, mass of body1, mass body2) and the distance between them, r. Goal is to estimate G!

Cavendish's 1798 data on mean density of the Earth, g/cm^3, relative to that of H_2O

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

Source: Cavendish [1798].

Sample mean $\overline{x} = 5.447931\ g/cm^3$

and sample SD = $s_X = 0.2209457\ g/cm^3$

Then the standard error for these data is:

$$SE(\overline{X}) = \frac{s_X}{\sqrt{n}} = \frac{0.2209457}{\sqrt{29}} = 0.04102858$$