UCLA STAT 110 A

Applied Probability & Statistics for Engineers

•**Instructor**: **Ivo Dinov,**

Asst. Prof. In Statistics and Neurology

•**Teaching Assistant:** **Christopher Barr**

University of California, Los Angeles, Fall 2004

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

Chapter 5

Joint Probability Distributions and Random Samples

5.1

Jointly Distributed Random Variables

Joint Probability Mass Function

Let X and Y be two discrete rv's defined on the sample space of an experiment. The *joint probability mass function* $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Let A be the set consisting of pairs of (x, y) values, then

$$P\left[(X,Y) \in A\right] = \sum_{(x,y) \in A} \sum p(x, y)$$

Marginal Probability Mass Functions

The *marginal probability mass functions* of X and Y, denoted $p_X(x)$ and $p_Y(y)$ are given by

$$p_X(x) = \sum_y p(x, y) \quad p_Y(y) = \sum_x p(x, y)$$

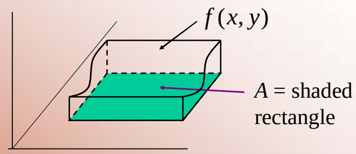
Joint Probability Density Function

Let X and Y be continuous rv's. Then $f(x, y)$ is a *joint probability density function* for X and Y if for any two-dimensional set A

$$P\left[(X,Y) \in A\right] = \iint_A f(x, y)dxdy$$

If A is the two-dimensional rectangle $\{(x, y): a \le x \le b, c \le y \le d\}$,

$$P\left[(X,Y) \in A\right] = \int_a^b \int_c^d f(x, y)dydx$$



$P\left[(X, Y) \in A\right]$

= Volume under density surface above A

Marginal Probability Density Functions

The *marginal probability density functions* of X and Y, denoted $f_X(x)$ and $f_Y(y)$, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \quad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx \quad \text{for } -\infty < y < \infty$$

Independent Random Variables

Two random variables X and Y are said to be *independent* if for every pair of x and y values

$$p(x, y) = p_X(x) \cdot p_Y(y)$$

when X and Y are discrete or

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

when X and Y are continuous. If the conditions are not satisfied for all (x, y) then X and Y are *dependent.*

More Than Two Random Variables

If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

If the variables are continuous, the joint pdf is the function f such that for any n intervals $[a_1, b_1]$, …, $[a_n, b_n]$, $P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n)$

$$= \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)dx_n \ldots dx_1$$

Independence – More Than Two Random Variables

The random variables $X_1, X_2, \ldots, X_n$ are *independent* if for every subset $X_{i_1}, X_{i_2}, \ldots, X_{i_n}$ of the variables, the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

Conditional Probability Function

Let X and Y be two continuous rv's with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x) > 0$, *the conditional probability density function of Y given that X = x* is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \quad -\infty < y < \infty$$

If X and Y are discrete, replacing pdf's by pmf's gives the *conditional probability mass function of Y when X = x*.

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

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

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

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

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

$$= \sum_R x f_{XY}(x,y)$$

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

$$= \sum_x \sum_{y \in R_x} (x-\mu_X)^2 f_{XY}(x,y) = \sum_{(x,y) \in R} (x-\mu_X)^2 f_{XY}(x,y)$$

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

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			Row Sum
		1	2	3	P{ ***X*** = i }
i	1	0.01	0.02	0.08	0.11
	2	0.01	0.02	0.08	0.11
	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 $f_X(x) > 0$

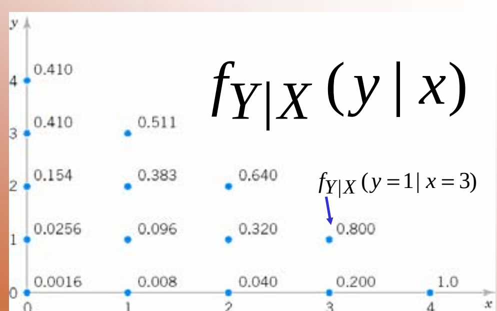
Conditional probability (Cont.)

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

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

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

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



$$f_{Y|X}(y \mid x)$$

$f_{Y|X}(y=1 \mid x=3)$

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

5.2

Expected Values, Covariance, and Correlation

Expected Value

Let *X* and *Y* be jointly distributed rv's with pmf *p*(x, y) or pdf *f* (x, y) according to whether the variables are discrete or continuous. Then the *expected value* of a function *h*(*X*, *Y*), denoted *E*[*h*(*X*, *Y*)] or $\mu_{h(X,Y)}$

is

$$= \begin{cases} \sum_x \sum_y h(x,y) \cdot p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)\,dx\,dy & \text{continuous} \end{cases}$$

Covariance

The *covariance* between two rv's *X* and *Y* is

$$\text{Cov}(X,Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right]$$

$$= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y)\,p(x,y) & \text{discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)\,f(x,y)\,dx\,dy & \text{continuous} \end{cases}$$

Short-cut Formula for Covariance

$$\boxed{\text{Cov}(X,Y) = E(XY) - \mu_X \cdot \mu_Y}$$

Correlation

The *correlation coefficient* of X and Y, denoted by Corr(X, Y), $\rho_{X,Y}$, or just ρ, is defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

Correlation Proposition

1. If a and c are either both positive or both negative, Corr($aX + b$, $cY + d$) = Corr(X, Y)
2. Corr(X, Y) = Corr(Y, X)
3. For any two rv's X and Y,

$$-1 \le \text{Corr}(X,Y) \le 1.$$

Correlation Proposition

1. If X and Y are independent, then $\rho = 0$, but $\rho = 0$ does not imply independence.
2. $\rho = 1$ or -1 iff $Y = aX + b$ for some numbers a and b with $a \neq 0$.

5.3

Statistics and their Distributions

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.

Random Samples

The rv's $X_1,\ldots,X_n$ 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.

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*.

5.4

The Distribution of the Sample Mean

Using the Sample Mean

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

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

$$2.\ V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$$

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

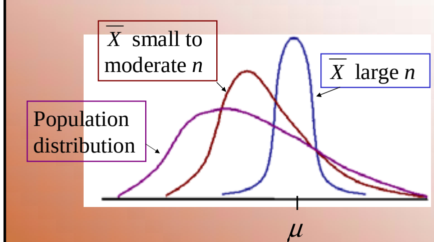
Normal Population Distribution

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

The Central Limit Theorem

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

The Central Limit Theorem



Rule of Thumb

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

Approximate Lognormal Distribution

Let $X_1, \ldots, X_n$ 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 $Y = X_1 X_2 \ldots X_n$ has approximately a lognormal distribution.

Central Limit Theorem – heuristic formulation

Central Limit Theorem:

When sampling from almost any distribution,

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

Show Sampling Distribution Simulation Applet:
file:///C:/Ivo.dir/UCLA_Classes/Winter2002/AdditionalInstructorAids/
SamplingDistributionApplet.html

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

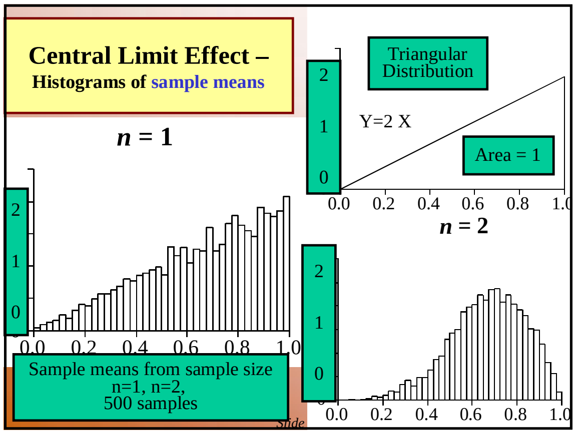
(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

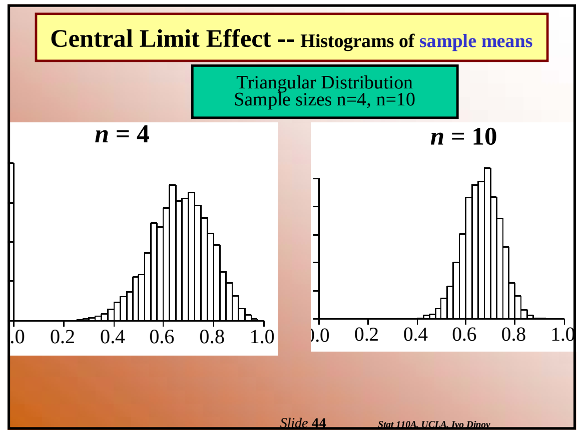
(3) $f_{X|y}(y) = f_X(x)$ for all x and y with $f_Y(y) > 0$

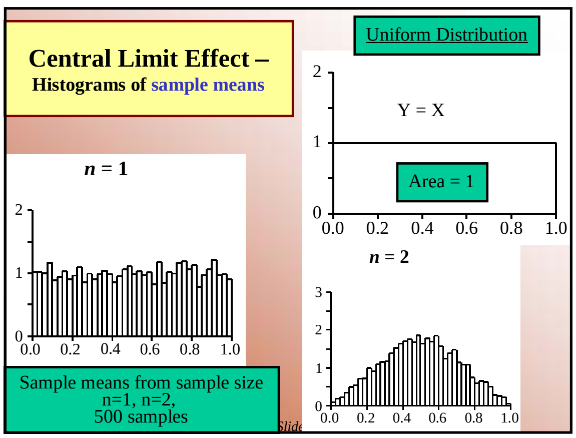
(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y respectively.

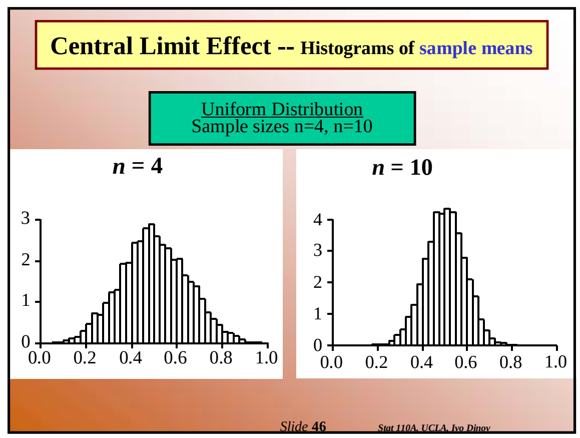
Recall we looked at the sampling distribution of $\overline{X}$

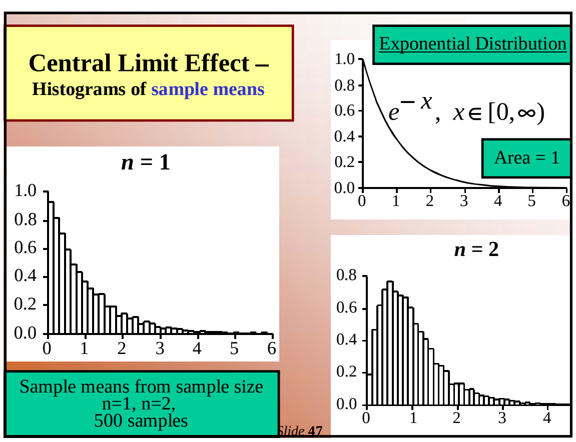
- For the sample mean calculated from a random sample, $E(\overline{X}) = \mu$ and $SD(\overline{X}) = \sigma/\sqrt{n}$, provided $\overline{X} = (X_1+X_2+ \ldots + X_n)/n$, and $X_k \sim N(\mu, \sigma)$. Then
- $\overline{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$. And variability from sample to sample 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.

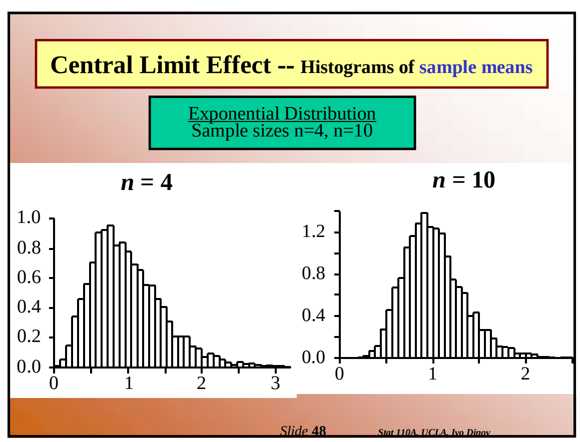

Central Limit Effect –
Histograms of sample means
n = 1
Triangular Distribution
Y=2 X
Area = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples


Central Limit Effect -- Histograms of sample means
Triangular Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 44
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

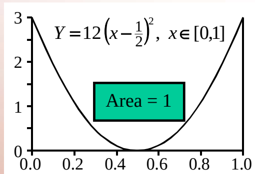

Central Limit Effect -- Histograms of sample means
Uniform Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 46
Stat 110A, UCLA, Ivo Dinov

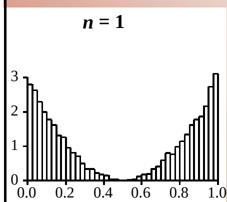

Central Limit Effect –
Histograms of sample means
Exponential Distribution
$e^{-x}, \; x \in [0, \infty)$
Area = 1
n = 1
n = 2
Sample means from sample size
n=1, n=2,
500 samples
Slide 47

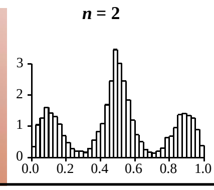

Central Limit Effect -- Histograms of sample means
Exponential Distribution
Sample sizes n=4, n=10
n = 4
n = 10
Slide 48
Stat 110A, UCLA, Ivo Dinov

Central Limit Effect – Histograms of sample means

Quadratic U Distribution



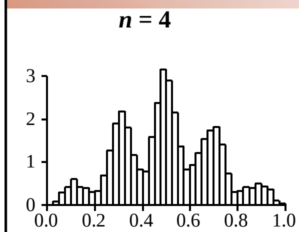


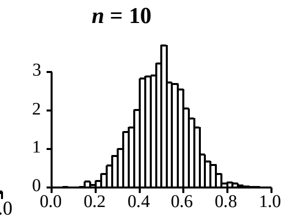


Sample means from sample size n=1, n=2, 500 samples

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,

$\overline{X}$ 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 $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, sample-avg,

Then $\overline{X}$ has a distribution which approaches N(μ, σ^2/n), as $n \to \infty$.

5.5

The Distribution of a Linear Combination

Linear Combination

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

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

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

Expected Value of a Linear Combination

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

Whether or not the X_i's are independent,

$$E(a_1X_1+\ldots+a_nX_n)=a_1E(X_1)+\ldots+a_nE(X_n)$$
$$=a_1\mu_1+\ldots+a_n\mu_n$$

Variance of a Linear Combination

If $X_1,\ldots, X_n$ are independent,

$$V(a_1X_1+\ldots+a_nX_n)=a_1^2V(X_1)+\ldots+a_n^2V(X_n)$$
$$=a_1^2\sigma_1^2+\ldots+a_n^2\sigma_n^2$$

and

$$\sigma_{a_1X_1+\ldots+a_nX_n}=\sqrt{a_1^2\sigma_1^2+\ldots+a_n^2\sigma_n^2}$$

Variance of a Linear Combination

For any $X_1,\ldots, X_n$,

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

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,

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

Difference Between Normal Random Variables

If $X_1, X_2,\ldots X_n$ 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.